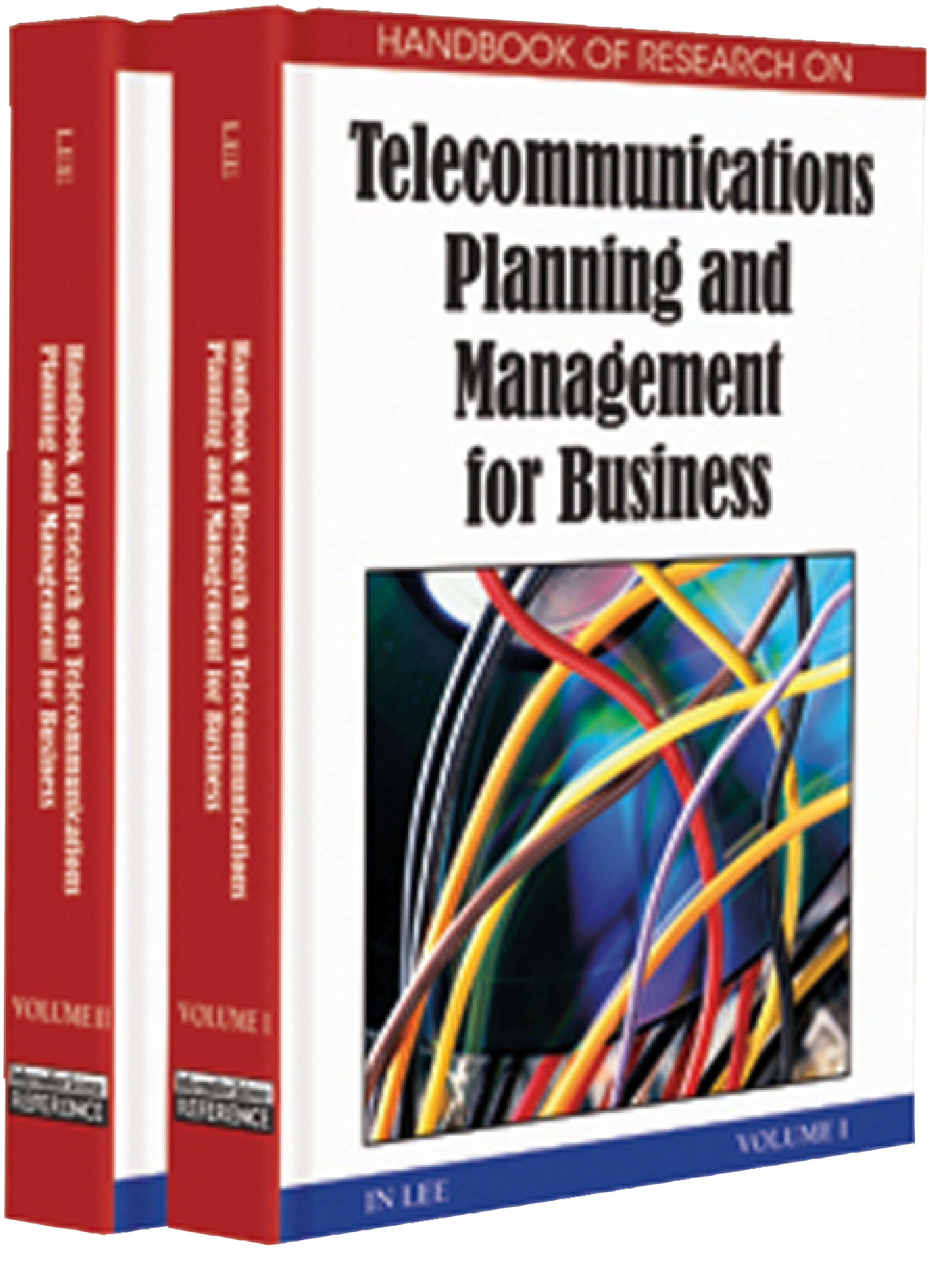

HANDBOOK OF RESEARCH ON
Telecommunications Planning and Management for Business
VOLUME I
IN LEE

Handbook of Research on Telecommunications Planning and Management for Business

In Lee
Western Illinois University, USA

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey • New York

Director of Editorial Content: Kristin Klinger
Senior Managing Editor: Jamie Snavely
Managing Editor: Jeff Ash
Assistant Managing Editor: Carole Coulson
Typesetter: Chris Hrobak
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on telecommunications planning and management for business / In Lee, editor.

p. cm.

Includes bibliographical references and index.

Summary: "This book provides original, in-depth, and innovative articles on telecommunications policy, management, and business applications"--Provided by publisher.

ISBN 978-1-60566-194-0 (hbk.) -- ISBN 978-1-60566-195-7 (ebook)

1. Telecommunication. 2. Telecommunication policy. 3. Telecommunication systems--Management. I. Lee, In, 1958-

HE7631.H35 2009

384.068--dc22

2008041546

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Handbook of Research on Telecommunications Planning and Management for Business is part of the IGI Global series named *Advances in E-Business Research (AEBR) Series*, ISBN: 1935-2700

Editorial Advisory Board

List of Reviewers

Jadranka Skorin-Kapov, *SUNY Stony Brook, USA*
Sanna Sundqvist, *Lappeenranta University of Technology, Finland*
Ching-I Teng, *Chang Gung University, Taiwan*
Nikolaos D. Tselikas, *National Technical University of Athens, Greece*
Belén Usero Sánchez, *Universidad Carlos III de Madrid, Spain*
Dimitris Varoutas, *University of Athens*, *Greece*
Sofie Verbrugge, *Ghent University/ IBBT, Belgium*
Xiaofei Wang, *Indiana University, USA*
Michael R. Ward, *The University of Texas at Arlington, USA*
Barney Warf, *University of Kansas, USA*
Chong Chin Wei, *Multimedia University, Malaysia*
Jason Whalley, *University of Strathclyde, UK*
James G. Williams, *University of Pittsburgh, USA*
Xiaotao Wu, *Columbia University, USA*
Kaiqi Xiong, *Texas A&M University Commerce, USA*
Kenneth C.C. Yang, *The University of Texas at El Paso, USA*
Sherali Zeadally, *University of the District of Columbia, USA*
Youlong Zhuang, *Columbia College, USA*

List of Contributors

Table of Contents

Section V
Telecommunications Technology and Applications

Detailed Table of Contents

in general and the market for broadband Internet access in particular. Both product and process (cost reducing) innovation is discussed. The chapter forms a compendium of available research on the intersection of telecommunications regulation and innovation.

Wireless/mobile technologies are absolutely necessary for any agent who wants to participate in the new information and communication technologies markets playfield. The mean for the deployment of the wireless/mobile technologies is the radio spectrum. A rational assignment and an efficient use of radio spectrum become *sine qua non* condition for the sector development. This chapter presents the reforms in the radioelectric spectrum management mechanisms that are currently being drafted (or that are even being applied) as well as to assess their advisability and timeliness. In particular, the chapter assesses the three deepest changes that are being considered: authorization of the secondary market, usage of auctions for primary allocation, and full liberalisation of spectrum usage.

The deployment of Next Generation Access networks (NGAs) is likely to have a significant impact on the telecommunications' value chain and, consequently, on the necessary regulatory remedies. In particular, unlike existing telephone networks, NGAs present regulators with a dilemma, in so far as the possibility of regulatory intervention after network deployment (negatively) affects investment incentives for such a deployment. This chapter reviews the current discussion surrounding NGAs and discusses some of the main regulatory challenges it presents.

The United States' wireless telephone industry has evolved from a minor segment of the communications industry to a major provider of voice, and increasingly data and video communication. The industry uses radio waves to transmit signals, and radio wave spectrum is regulated by the federal government. Moreover, local transmission requires unobstructed antennae, which in rural and suburban areas has led to the construction of wireless towers. States and municipalities have sought to regulate the construction of such towers, citing issues of aesthetics and health. The development of the wireless industry has been constrained by such government regulation. This chapter discusses the impact of government on the market structure of the wireless industry.

Since 1986, the Internet has developed into a global network enabling users worldwide to connect to each other to exchange information and data. This chapter focuses on current interconnection settlement arrangement models that disfavor ISPs and end-users in the Asia-Pacific region. After reviewing the Internet market and digital divide in the region, the chapter summarizes the main current IIS issues between the Asia-Pacific and Western regions into three categories of concern: inequity, anticompetitive practices and the threat of the "balkanization" of the Internet. Practical recommendations to resolve these issues and improve the Asia-Pacific IIS market are discussed from regional and international perspectives.

This chapter examines cross-national data in order to identify possible factors related to the observable patterns of Mobile Virtual Network Operator (MVNO) penetration. The study conducts an economic assessment of market structure and environment for different countries' MVNO penetration. The data on market structure is analyzed by means of cluster/factor analysis techniques in order to group countries according to their market environments. Then, regression equation analysis is used to investigate the relations of MVNO penetration and independent variables. The results show that MVNO penetration significantly relates to market structure and performance. These results also imply an appropriate policy to promote MVNO diffusion.

This chapter focuses on the internationalization processes of national telecommunications companies (telcos) from small and open economies (SMOPECs) who have moved from a domestic monopoly to an actor within the global industry. This chapter aims to increase our understanding of *how* these companies have internationalized, *what factors* have been the most influential in this process, and *how the position* of these companies *has changed* in the evolving value network of the industry. Case examples are used to illustrate the internationalization processes of telcos from SMOPECs within the context of the whole industry value network.

This chapter analyzes the role of non-market actions (NMAs) in the competitive position of firms in the European mobile telecommunications industry. First, the bases and uses of NMAs are reviewed and classified. Second, the use of legal actions as NMAs to erode the competitive position of first-mover firms is examined. Third, the findings of an empirical study confirming the effectiveness of legal actions as competition tools in the European mobile telecommunications industry are analyzed.

Fiber optics forms the pivotal telecommunications technology of the contemporary global economy, offering greater speed and security than other modes. This chapter reviews the historical development of this communications technique. It then views its rapid growth within the context of contemporary globalization. Fiber's role in contemporary urban restructuring is noted. The chapter then turns to the spatial distribution of the world's fiber lines, noting major transatlantic and transpacific markets and newer systems.

This chapter categorizes three classes of typical services and five industrial participants with regard to Korea's home network industry. The chapter examines the government plans on home network services in the early stage of evolution, consumer value expectations and current home network service levels. Then, the chapter investigates business interests of the industrial participants comprising convergent home network services, and explores the causes of conflicts among them. The chapter offered policy suggestions from digital convergence perspective to help removing the obstacles that hinder business cooperation among industrial participants of the home network industry.

This chapter discusses managerial implications and strategic alliance opportunities of the home network service providers in Korea. Based on in-depth analysis of recent industrial status of home network services in Korea, the chapter first analyzes business models of each interested parties participating converging services of home network industry. Next, in order to complement the weakness of participant's business models, strategic business alliance opportunities and managerial implications are drawn upon resource-based view (RBV) theories. Subsequently, strategic propositions for the success of the home network are suggested at the firm level, with respect to three main converging home network services.

This chapter explains what organizational culture is and analyzes its importance for the management of any company. Organizational culture must not be ignored during the decision-making process and managers must understand the existing culture of their organization in order to achieve their targets and to meet their goals. This chapter presents the theoretical link between organizational culture and a variety of variables, which affect organizational performance and efficiency, directly and indirectly. Such variables are knowledge management, organizational climate, leadership, quality, innovation and entrepreneurship, human resource management, and employee behavior.

This chapter analyzes the importance of organizational culture through a set of strategies and enterprising practices in order to emphasize the culture's necessity for the effective management of any organization and especially for the management of telecommunication companies. The way culture affects the success and failure of a series of enterprising practices, such as mergers, acquisitions, strategic alliances and joint ventures is presented. All these create a rather interesting framework in which organizational managers should cultivate and implement the appropriate organizational culture in order to keep their organization competitive and well managed. Finally, case studies from the telecommunications industry will be presented, a field where culture plays a dominant role in either changing procedure or in developing new enterprising practices.

This chapter discusses the adoption of mobile services. Not all mobile services experience equal interest among end-users. Furthermore, not all interest towards the service leads to actual use of the service. This chapter studies the dynamics of mobile service adoption in a cross-service setting both qualitatively and quantitatively. The study explores what drives the intention and actual use of some of the emerging mobile services, thus contributing to both industry and academic readers.

This chapter familiarizes the reader with the fundamentals of communication protocols in a special application area called telecommunications management. It presents the most widely used protocol, the Simple Network Management Protocol (SNMP), the open system management protocols, and the mobile internet management protocols for Authentication, Authorization and Accounting (AAA). Towards the end of the chapter, the chapter also provides tables to help the reader compare the management protocols presented.

This chapter presents the characteristics and the architecture of each generation of cellular networks (1G, 2G, 3G and 4G). Then, it exposes different planning problems related to each generation followed by a short description of different solutions that have been proposed in the literature.

Volume II

IP network connectivity is becoming so important for business that it is being compared today, to basic utilities such as water and electricity. Today, more than any time before, companies are realizing the need to adopt and implement a comprehensive IP Address Management (IPAM) strategy. This chapter turns the spotlight on the most important challenges in IPAM and attempts to address some of the solutions and best practices to tackle these challenges. The chapter underlines the need for organizations to adopt proven IPAM best practices and deploy good automated IPAM tools. This will put them in a better position to expand and leverage their existing networks, while optimizing their IP address space in a secured and controlled fashion.

This chapter focuses on digital cable TV networks as a convergent network with telecommunications networks and the Internet that provides broadcasting TV and radio, telecommunications services, and

IP-based publishing and E-Commerce. The chapter first traces the technological evolution of cable TV, highlighting recent developments in digitalization and convergence. The transformation of cable TV networks from channel operators to unified platforms is discussed. Furthermore, this chapter examines the value chain and collaborative opportunities among the participants in the digital cable TV revolution.

The diffusion of a Web 2.0 product or services is, unlike to traditional consumer or industrial goods, not only based on purchase. Full acceptance of Web 2.0 platforms occurs by recurring utilization. This chapter focuses on diffusion characteristics of this innovative category of ICT products and provides management concepts for competition. The concept of critical mass is applied to different growth scenarios. Additional success factors are discussed. Markets are never settled, due to the ever changing and oscillating conditions. The chapter shows that there is always a chance to capture a market or at least to grow against competition in a Web 2.0 setting.

This chapter discusses the potential of WiMax technology to deliver personal broadband and fixed Internet capabilities. The chapter applies components from the global diffusion of the Internet framework to analyze the current state of WiMax deployment utilizing the following six dimensions: pervasiveness, geographic dispersion, sectoral absorption, connectivity infrastructure, organizational Infrastructure, and sophistication of use.

VoIP is the use of Internet protocols to provide telephone services that have previously been delivered over traditional telephone networks. Advantages of VoIP include cost, portability, and functionality which are the main reasons that many consumers and small businesses are considering this technology as a replacement to traditional telephone services. There are however risks associated with VoIP services which impact quality and security of the phone system for voice communications. This chapter reviews issues related to quality and security faced by consumers and small businesses. Recommendations are provided to improve call quality and mitigate threats faced in the VoIP environment.

The deployment of VANETs for E-Business is rapidly approaching, and their success and safety will depend on viable security solutions acceptable to consumers, manufacturers and governments. This chapter presents the emerging security issues in Vehicular Ad hoc Networks (VANETs) for E-Business along with some of the solutions. Security issues discussed are authentication, trust, confidentiality, non-repudiation and detection of malicious users.

Network planning and management in a large organization like a university is a complex task. A large university is an inherently demanding environment in terms of telecommunications services offered and technologies used. This chapter presents and discusses network management and planning issues in the Aristotle University of Thessaloniki, Greece. Examples of network management procedures are given. It presents the data model for service provision. The network expansion and cost reduction case studies are discussed. A SWOT analysis about University's Telecommunications Center is given. Finally, it discusses the migration analysis for future upgrades that will fully enable the use of emerging technologies such as Voice over IP (VoIP).

This chapter discusses the applications, state-of-the-art technologies, planning methods and business models for wireless community networks, and provides an integrated presentation of these essential parts with examples. It gives an overview about the state-of-the-affairs of wireless community networks presenting the driving forces, stakeholders, and services built upon applications in some carefully selected projects. Then, it discusses the technological aspects and suggests a design methodology for planning wireless community networks. The application of this methodology is illustrated via an ongoing digital city project in Hungary.

This chapter addresses an area of science, human factors, with particular relevance to developing and deploying business mobile systems that are fit for purpose. The value of human factors is discussed, before moving onto some general guidelines to ensure that a mobile technology is suitable for its users. It is critical to take an approach that puts users at the heart of requirements, build and deployment of a new technology or service. This approach is presented as a three-stage model of context analysis, specification and design, and evaluation.

IP networks are the federative transport networks for a large set of emerging services. These services demand hard guarantees in term of the service availability, experienced Quality of Service (QoS) and robustness. In order to meet QoS requirements of these services in an inter-domain context, several issues should be solved. This chapter focuses on two issues: provider-to-provider agreements and enhancements to inter-domain routing protocol to convey QoS-related information. This chapter provides a framework suitable for the promotion of QoS-enabled services with an inter-domain scope: the Parallel Internet. This concept is a viable way for the management of IP resources so as to deliver end-to-end QoS-enabled services.

This chapter proposes technical playout control management and planning strategies and compares the strategies IP Telephony Service Providers can choose with the aim of saving money and offering a better quality of service. The chapter also introduces the state-of-the-art quality index for IP Telephony and provides the reader with examples on some economic scenarios of IP Telephony.

Most of the existing networks (e.g., telecommunications, industrial control, enterprise networks etc.) have been globally connected to open computer networks (Internet) in order to decentralize planning, management and controls in business. Most of these networks were originally designed without security considerations, thereby making them vulnerable to cyber attacks. This has given rise to the need for efficient and scalable intrusion detection systems (IDSs) and intrusion prevention systems (IPSs) to secure existing networks. It has been proven that the right combination of security techniques always protects networks better. The novelty of our proposed IDS technique presented in this chapter lies in its efficiency and ability to eliminate most of the limitations of existing IDSs and IPSs, thereby ensuring high level network protection.

This chapter presents high-speed networking technologies and standards such as ATM, Fast Ethernet, 10 Gigabit Ethernet, SONET, RPR, PBT, PBB, T-MPLS and Optical Transport Network (OTN). It considers the requirements imposed to high-speed networks by multimedia applications and analyses crucial issues of high-speed networking such as bandwidth problems, discarding policies and fast broadcast.

This chapter introduces the novel concept of Cognitive Radio for wireless telecommunications. Cognitive radios are a new type of radio devices that include cognition and reconfigurability features. This chapter presents the reader the new paradigm that cognitive radio networks have created. It focuses on the challenges in PHY layer and MAC sublayer and the most relevant studies in these fields.

This chapter addresses the role of Game Theory for wireless communications. It introduces Game Theory and its fundamentals, and gives an overview of the applications of Game Theory in the wireless networks for PHY layer and MAC sublayer.

Mobile commerce is an emerging manifestation of electronic commerce that bridges the domains of Internet, mobile computing and wireless telecommunications in order to provide an array of sophisticated

Service and content delivery over the Internet is currently supported by overlay architectures of different types. There is a trend towards distributed computing and service creation in peer-to-peer and grid networks, which are able to overcome performance bottlenecks of client server architectures. Overlays are deployed for single applications or as multi purpose infrastructure for communities with focus on their special demands. Various overlay structures have also developed on lower network layers. The motivation for those overlays is to bridge or extend one networking technology on top of another in order to build common widespread platforms through heterogeneous telecommunication environments. This chapter addresses such approaches especially within the standardization of Internet Protocols (IP), where the main focus is on evolving techniques on higher layers.

Distributed Hash Tables (DHTs) have been used in Peer-to-Peer networks to provide query lookups in typically O(log n) messages whilst requiring maintenance of only small amounts of routing state. This chapter proposes ROME, a layer which runs on top of the Chord DHT to provide control over network size through monitoring of node workload and proposes the use of processes to control the addition or removal of nodes from the network. This chapter shows that this technique can reduce further the hop counts in networks where available node capacity exceeds workload, without the need to modify any processes of the underlying Chord protocol.

The network carrier must utilize the full potential of existing physical infrastructure before any new investments can be considered. Much work has been done to develop optimal design methodologies that generate a "best" design but designs often have to be changed to meet operational considerations, potentially mitigating any optimization benefits. This chapter proposes to expand the optimal network design process by applying a process borrowed from Finance and Operations Research/Management Science literature known as the Efficient Frontier or production frontier analysis. By comparing any subsequent design to the frontier, the network designer can have an understanding of the impact of changes in the design strategy to the long-term cost effectiveness.

Foreword

INTRODUCTION

For 30 years I taught graduate courses, consulted for many companies and governments around the world, and did research in the area of information systems and telecommunications as well as spending several years as the CIO of a telecommunications company. I wish I had this book to provide the valuable data, information, knowledge, recommendations, experiences and guidance contained herein for all those activities.

We live in the age of communications and the policies, regulations, technologies, services, economics, infrastructure, planning, design and end user demands throughout the worldwide telecommunications industry produce forces that are now pushing the industry forward at a rapid pace. The end user's demand for more bandwidth; increased reliance on mobility services; desire for better security and privacy, and the end users' assessment of cost versus performance are driving the providers, regulators, researchers, and policy makers to reassess all aspects.

Thus far, the industry has reacted to these forces driving major infrastructure and technological change through consolidation, as carriers merge and equipment suppliers struggle to survive. Both on the domestic and international levels, regulators and government agencies are struggling to cope with rapid, profound changes in an industry that had been predictable for more than 100 years.

The technologies underlying the industry's changing infrastructure can be summed up as: IP transport's ability to merge voice, video and data; the long-haul fiber glut and the resulting reduction in the cost of bandwidth; the wide variety of local broadband wired and wireless solutions; and increasing security requirements; especially the ability to protect users on the Internet. The telecommunication's industry is riding the crest of worldwide demand for new video-enabled services. The emergence of video as a major force is reshaping mobile and fixed line telecommunications by the investment of large sums of money into new infrastructure which will cause a new telecommunication's landscape to be shaped. This new landscape will be formulated around the following 6 components:

1. Telecommunications Policy
2. Telecommunications Industry
3. Telecommunications Business Management
4. Telecommunications Technology Management
5. Telecommunications Technology and Applications
6. Telecommunications Network Design

This handbook presents empirically-based evidence that provides valuable information and knowledge along with recommendations for the participants in each of these components.

TELECOMMUNICATIONS POLICY

For policy makers, regulators, and legislators around the world, this book provides invaluable data and recommendations that will help them deal with policies related to the tremendous growth in telecommunications and the demand for new, high performance services. Regulators in many countries are finding that existing regulatory frameworks are not suitable for dealing with these new services and related technologies. A framework in this section of the book is presented that is comprised of five major components: consumer and intellectual property protection, market and resources access, and environmental protection. New and innovative approaches are discussed in many chapters related to these issues. Since wireless/mobile technologies are currently at the center of many policy and regulatory issues suggested reforms in the radioelectric spectrum management mechanisms and rational assignment and use of the radion spectrum are discussed in several chapters of this section. The role of regulation on incentives to invest in telecommunications in Europe, Australia, and the United States, especially in Next Generation Access Networks (NGA), is presented in both quantative as well as qualitative terms. Policy makers and regulators around the world will find this section highly useful.

TELECOMMUNICATIONS INDUSTRY

The telecommunications industry is undergoing changes that have an effect on revenues, costs, market share, profitibality, competition, planning, efficiency, operations, personnel, and management practices. This section of the book provides extremely useful data, information, knowledge, and innovative insights into these topics. The convergence of IT and telecommunications and the emergence of new technologies and new entrants with new business models are discussed in several chapters of this section of the book. The local exchange telephone market is no longer considered to be a natural monopoly as a result of technological advancements. Digital technology has enabled traditional telecommunications companies to transition their network architecture from one based on copper wires and circuit switches for providing high quality transmission of point-to-point voice signals to a multi-faceted, general network based on fiber optic cables and packet switches capable of providing an array of voice, data, and video services. As noted in one chapter, convergence is fundamentally changing the nature of what it means to be a telecommunications company. One example of the impact of convergence presented in this section is what is happening in Korea. In Korea, the home network industry has been the object of great expectations among the government and business community. Thus far, however, it has proved unable to fulfill these expectations. On the basis of analyses, this chapter offers policy suggestions from a digital convergence perspective to help remove the obstacles to business cooperation among the participant industry sectors of the home network industry. Managerial implications for the participating firms are discussed in terms of adopting a new business model framework. The Internet has developed into a global network enabling users worldwide to connect to each other to exchange information and data. The initial peering arrangements have generally been replaced by commercial transit arrangements, as backbone providers seek to recoup their network infrastructure investments and generate commercial profits. This is a key cause of the issues and debates that have emerged between developed and developing countries about international Internet interconnection services (IIS). This book presents data on current interconnection settlement arrangement models that disfavor ISPs and end-users in the Asia-Pacific region and summarizes the main current IIS issues between the Asia-Pacific and Western regions. As part of the industry analysis the book presents a techno-economic analysis of Mobile Virtual Network Operators by examining cross-national data in order to identify possible factors related to the observable patterns of Mobile Virtual Network Operator (MVNO) penetration including an economic assessment of market structure and environment for different countries' MVNO penetration. Fiber optics forms one of the industry's pivotal telecommunications technologies of the global economy, offering greater speed and security than other modes. Fiber's role in contemporary urban restructuring is discussed. One chapter in the book analyzes and discusses the spatial distribution of the world's fiber lines, noting major transatlantic and transpacific markets and newer systems. The enormous construction boom of the 1990s and early 2000s which led to severe overcapacity, with significant economic fallout is presented. One chapter in the book analyzes the role of non-market actions (NMA's) in firms' competitive position within the European mobile telecommunications industry. The bases of non-market actions are reviewed and classified as well as the use of non-market actions by the mobile phone operators is described using a particular case to describe the relationship between the use of legal actions and the first mover advantage in the European mobile telecommunications industry.

TELECOMMUNICATIONS BUSINESS MANAGEMENT

Having taught courses and performed research in the area of telecommunications and information systems for a major university for over 30 years, it was quite a transition when I became the Chief Information Officer (CIO) for a Telecommunications company and discovered the "business" of telecommunications. This book provides a number of extremely useful chapters on the business management aspects of telecommunications which I wish I had read before having to learn it on the job. Telecommunications is a business that has many aspects that traditional business and planning models do not encompass. One example of trying to deal with some of the idiosyncratic aspects of the telecommunication's business is the chapter that examines an implementation of an Activity-based Costing (ABC) system in a Portuguese telecommunications firm called Marconi. It is argued that they changed its management accounting system due to efficiency and institutional pressures from its constituencies, namely the Portuguese telecommunications regulators, its managers, and its competitors following the liberalization of the Portuguese and European Union telecommunications market. A chapter examines business process innovation which involves creating systems and methods for improving organizational performance and the interesting and innovative topic of intelligent networking which represents the combination of software, technology and electronic pathways that makes business process innovation possible for both large and small organizations based on the central tenet that an intelligent network is not one network, but a series of networks designed to enhance worldwide communication for business and residential users. Other chapters focus on the internationalization processes of national telecommunications companies (telcos) which have moved from a domestic monopoly to a player within the global industry. Every organization has a culture which guides its policies and management principles and can not be ignored. Several chapters in the book present data on the importance of organizational culture for the management of any organization, the theoretical link between organizational culture, and a variety of variables which affect organizational performance and efficiency, directly and indirectly. Such variables are knowledge management, organizational climate, leadership, quality, innovation and entrepreneurship, human resource management, and employee behavior. Planning is essential for any organization and is especially important in a fast changing industry such as telecommunications. Planning and operational management models are presented in several chapters of this section of the book which encompass business strategies, knowledge management, customer satisfaction, customer demand, churn reduction, cost-benefit analysis, investment analysis, upgrade strategies, adoption of new technologies (such as VoIP), convergent technologies, and human factors. These are discussed in the context of various countries, technologies, and business domains including academia.

TELECOMMUNICATIONS TECHNOLOGY MANAGEMENT

Telecommunications is functionally dependent on a wide array of technologies. These technologies have been changing at a rapid pace thereby producing an array of technological issues that have placed a greater emphasis on managing these technologies. This book presents empirical data as well as theory-based accounts related to managing the many issues created by these new technologies. For example, the diffusion of a Web products or services is unlike traditional consumer or industrial goods but is characterized by recurring utilization and the permanent supervision of a platform regarding its compliance with qualitative, as well as ethical and legal standards. Web Services technologies enable the proper communication of processes deployed in quite hostile environments such as the Internet. Cellular networks have experienced significant changes over the last few years. In order to keep up with this constant evolution, planning tools must also adapt themselves in order to reflect the particularities and architecture of each generation. Voice over IP or VoIP is the use of Internet protocols to provide telephone services that have previously been delivered over traditional telephone networks. Advantages of VoIP include cost, portability, and functionality as a replacement to traditional telephone services but as pointed out in several chapters there are risks associated with VoIP services which impact quality and security of the phone system for voice communications. The chapter on the basics of technology management focuses on management functions, the managed objects, the management information base, and the management protocols. The fundamentals of communication management protocols have changed with the most widely used protocol, the Simple Network Management Protocol (SNMP) but new open system management protocols and the mobile internet management protocols are used for Authentication, Authorization and Accounting (AAA). One chapter in this section points out that IP network connectivity is becoming as important for businesses as basic utilities. IP addresses are among the most important resources to be managed. Ready access and flawless handling of IP address usage, assignment, tracking and reallocation help in enhancing network reliability and security, while enabling more efficient network expansion and troubleshooting. The chapter on managing IP Addresses is the best one on this subject that I have ever read. The chapter, "Planning of Wireless Community Networks", discusses the applications, state-of-the-art technologies, planning methods and business models for wireless community networks and provides an integrated presentation of these essential parts with examples which is invaluable for managing a wireless network.

TELECOMMUNICATIONS TECHNOLOGY AND APPLICATIONS

Technologies are useless until utilized for some application. Telecommunication organizations invest large sums of money in technologies with the assumption that they will receive a return on that investment based on how it is applied to provide services that generate revenue and/or reduce costs. But applying technologies for such purposes is filled with problems. The section of the book "Telecommunications Technology and Applications" discusses in a rather complete manner the problems and issues faced by providers in using technologies and especially new technologies for new and improved applications. The utilization of service and content delivery over the Internet is currently supported by overlay architectures of different types. There is a chapter which focuses on two issues related to IP, federated networks, namely: provider-to-provider agreements and enhancements to inter-domain routing protocols to convey QoS-related (Quality of Service-related) information. A concept called Meta-QoS-Class is introduced together with an enriched version of the Border Gateway Protocol. The chapter on the use of Web Services for management describes how Web Services allow the integration of low-level activities (e.g., retrieving monitoring information from gateways) with high-level business processes (e.g., creating a new product and its marketing strategy.). Despite clear advantages, Web Services-based management does not come for free; since Web Services are based on XML documents, its performance, compared with traditional management technologies, may represent an important drawback. The chapter on intrusion detection notes that most existing networks have been globally connected to open computer networks (Internet) in order to decentralize planning, management and controls in business but most of these networks were originally designed without security considerations, thereby making them vulnerable to cyber attacks. This has given rise to the need for efficient and scalable intrusion detection systems (IDSs) and intrusion prevention systems (IPSs) to secure existing networks which are discussed in detail. Fault management is presented as systems that are based on the analysis of Key Performance Indicators (KPIs) in order to identify the cause of network malfunctions. The chapters on m-based (Mobile based) commerce and pervasive computing and the use of agent technologies reveals many hidden issues in various applications. The discussion of a novel concept called Cognitive Radio for wireless telecommunications is fascinating. Cognitive radios are a new type of radio device that include cognition and reconfigurability features, thus creating a new paradigm for wireless environments. A chapter also describes the creative use of game theory for PHY layer and MAC sub-layer in Wireless Telecommunications. The chapter on high-speed multimedia networks: critical issues and trends is a very complete coverage of the range of technologies and standards ranging from Asynchronous Transfer Mode (ATM) and Fast Ethernet to Multi Protocol Label Switching (T-MPLS) and Optical Transport Network (OTN). Chapters on Open API and protocols as well as security concerns are well worth reading. The chapter on the role of telecommunications in Precision Agriculture illustrates how an application typically requires the convergence of several technologies to make the application a success. Since most farms are by definition rural and do not have high speed access to the Internet, the problem of moving 10's, 100's and even 1000's of megabytes of essential information and data must be solved to help make precision agriculture easy to use and even transparent to the grower. Wireless technology can fill this gap and help make precision agriculture the standard in the future.

TELECOMMUNICATIONS NETWORK DESIGN

Having taught courses on network design and having been involved in designing several networks including a commercial DSL network, I could have benefitted from the chapters in this section of the book. In designing the DSL network, we had to utilize existing network facilities of an incumbent wireline company. Several chapters in this section emphasize that the network carrier must utilize the full potential of existing physical infrastructure before any new investments can be considered. There is a trend towards distributed computing and service creation in peer-to-peer and grid networks, which are able to overcome performance bottlenecks of client server architectures but these are not without issues that must be planned for and tested. One chapter focuses on planning and traffic engineering for link bandwidth and buffers as main resources in communication platforms based on measurement and statistical properties of traffic growth and variability which are critical for managing these technologies. Much work has been done to develop optimal design methodologies that generate a “best” design but designs often have to be changed to meet operational considerations, potentially mitigating any optimization benefits. One chapter in this section proposes to expand the optimal network design process by applying a process borrowed from Finance and Operations Research/Management Science literature known as the “Efficient Frontier” or production frontier analysis. Using this idea, a portfolio of designs differing in cost, and number, size and location of nodes will be developed and from that the efficient frontier defined. By comparing any subsequent design to the frontier, the network designer can have an understanding of the impact of changes in the design strategy to the long-term cost effectiveness. As one chapter notes; the cellular principle is an effective way to guarantee efficient utilization of the offered radio band and although PCS networks use the cellular principle, the next generation of PCS networks needs more improvements in location management to face the increased number of users. Both an Enhanced Profile-Based Strategy (EPBS) for small-scale roaming and a Caching Two-Level Forwarding Pointer (C2LFP) strategy for large-scale roaming are introduced. An evolutionary method, using a constraint Genetic Algorithm (GA) has been handled to achieve network parameters optimization and is presented in a case study format. A chapter in this section presents optimizing P2P Networks through adaptive overlay construction. Another chapter in this section presents a framework for solving a security-aware resource optimization problem under the SLA constraints of .trustworthiness, percentile response time and availability using an efficient numerical procedure. The chapter on “Trends, Optimization and Management of Optical Networks” reviews optical networking, its current issues, trends, and ways to efficiently design and manage a wide range of optical networks. Efficient management and utilization of geographically different networks (such as wide area networks, metro networks, and access networks) present different optimization problems dealing with the placement of special optical equipment whose solutions are summarized in terms of current algorithmic approaches and suggest possibilities for further research.

SUMMARY

I highly recommend this book for professors teaching courses and performing research in network management, network analysis and design, network protocols, network regulations and standards, wireless networks, network performance metrics, telecommunication services, and network security. I would also recommend it to Masters and PhD students looking for ideas for a thesis.

For those working in the telecommunication industry whether it is at the operational level, managerial level or executive level, this book can provide solutions and approaches to existing problems as well as generate new approaches to services, control, and measuring performance. Likewise, anyone involved in networking at almost any level can benefit from one or more chapters of this book to provide a more responsive, efficient, secure, and service-oriented telecommunications network.

This book is a must for policy makers and regulators around the world starting the FCC in the United States to the appropriate bodies in Australia, United Kingdom, European Union nations, countries in the Far East such as China, and countries in the Middle East. The empirical data, creative analysis, and innovative recommendations for policy and regulation of this rapidly changing and high impact area affecting the social and political aspects of every country in the world are captured in this book.

Preface

Over the last decade the telecommunications industry has experienced rapid transformation. Many factors, such as digital convergence, wireless networks, deregulation, and globalization, have contributed to this transformation. Convergence across telecommunications and the Internet has caused major upheavals in the value chain of the telecommunications industry. Telecommunications convergence coupled with deregulation has created numerous opportunities and challenges for both incumbent telecom operators and new entrants. The convergence has put telephone companies in a fierce competition with cable TV operators for voice, data, and video services. Telephone companies are transforming their provision of voice and other broadband services to the Internet layered fiber-optic networks to embrace the IP-based telecommunications convergence. While some telecom operators build new networks, other operators offer services with networks of incumbent operators. The competition among these incumbents and new entrants will become fiercer in the future, and telecommunications planning and management become more important for telecom managers.

Among stakeholders, governments had a major impact on the industry's structure and competition through policies on licensing, mergers/acquisitions, and other regulations. Policy makers face challenges of facilitating competition into quasi-monopolistic markets, and at the same time, developing favourable conditions for the capital investments needed to sustain and improve the network infrastructure (Bauer and Bohlin, 2008). For emerging industries such as the mobile services industry, regulators have to deal with a wide variety of stakeholders. Further, as mobile technology operates in a dynamic environment and undergoes rapid changes, regulatory definitions become a moving target which implies that regulators should constantly acquire industry-specific knowledge over time (Tallberg et al., 2007; Ubacht, 2004). Consequently, regulation for the mobile industry becomes an extremely complex and involved undertaking for all stakeholders.

One of the most significant developments in the telecommunication technologies of the past decade is wireless networks. The wireless network industry is evolving at a very fast pace. The diffusion of mobile handsets and services has achieved spectacular growth in many countries around the world. WiMax and Wi-Fi have shown the potential to be a leading technology in the wireless industry for the service providers. While the third generation networks are deployed, the next generation networks are developing. The key feature of the next generation is an integration of different access networks (HSDPA, UMTS, LTE, GSM, Wi-Fi, Wi-Max, etc.) around a common IP.

The telecommunication market is shifting from a stable market to an increasingly user-driven market place. The telecom operators must have the ability to rapidly develop, deploy, and manage services to meet customers' dynamically changing needs. The success of a telecom operator depends on the operator's ability to understand and manage the organizational variables such as knowledge management, organizational climate, leadership, service quality, innovation and entrepreneurship, human resource management, and employee attitude.

According to the U.S. Bureau of Labor Statistics (BLS), jobs in telecommunications are projected to grow faster than other occupations in the coming years. The incredible growth of telecommunications technology and the demand for a new generation of telecommunications employees, managers, and leaders facilitated the offering of Telecommunications Management programs in many higher education institutions in the United States and around the world. Although new technologies and models are continually appearing, it is high time to take stock of the new knowledge in telecommunications planning and management. The *Handbook of Research on Telecommunications Planning and Management for Business* provides a repository for government policy-makers, researchers, and industry practitioners to share and exchange their policy and research ideas, theories, practical experiences, discuss challenges and opportunities, and present tools and techniques in all aspects of telecommunications planning and management.

This book is composed of 60 chapters and divided into 6 segments: Section I, Telecommunications Policy, which discusses various issues and recommendations for policy-makers and regulators in the telecommunications industry; Section II, Telecommunications Industry, which addresses competition, dynamics, and trends in the industry; Section III, Telecommunications Business Management, which discusses various business management topics such as organizational culture, management methods, customer management, investment, and practices; Section IV, Telecommunications Technology Management, which presents various technology management methods and techniques for security, protocol management, and new technology services; Section V, Telecommunications Technology and Applications, which addresses the state-of-the-art technologies, protocols, Web services, and implementation issues faced by telecom service providers; and Section VI, Telecommunications Network Design, which discusses methods, issues, and new trends in the network design . A brief introduction of each chapter follows.

Section I: Telecommunications Policy, consists of 6 chapters. Chapter I, "Static and Dynamic Efficiency in the European Telecommunications Market: The Role of Regulation on the Incentives to Invest and the Ladder of Investment" by Walter Distaso, Imperial College (UK), Paolo Lupi, AGCOM (Italy), and, Fabio Manenti, University of Padua (Italy), provides evidence of the effectiveness of European national regulatory authorities in applying the basic principles of the ladder of investment. The analysis discusses and compares the regulatory approach adopted in 12 European countries from January 2005 to July 2007. Chapter II, "Is Regulation a Roadblock on the Information Highway?", by James E. Prieger and Daniel Heil, Pepperdine University (USA), covers the theoretical and empirical impacts regulation has on innovation. It presents empirical evidence that regulation dampens firms' incentive to innovate in the telecommunications industry in general and the market for broadband Internet access in particular. Both product and process (cost reducing) innovations are discussed. Chapter III, "Reforms in Spectrum Management Policy" by Claudio Feijóo, Universidad Politécnica de Madrid (Spain); José Luis Gómez-Barroso and Asunción Mochón, Universidad Nacional de Educación a Distancia (Spain), presents the reforms in the radioelectric spectrum management mechanisms that are currently being drafted (or that are even being applied). In particular, the chapter assesses three major changes that are being considered: authorisation of the secondary market, usage of auctions for primary allocation, and full liberalisation of spectrum usage. Chapter IV, "Next Generation Access Networks and their Regulatory Implications" by Ricardo Gonçalves and Álvaro Nascimento, Universidade Católica Portuguesa (Porto) (Portugal), reviews the current discussion surrounding Next Generation Access networks (NGAs) and discusses some of the main regulatory challenges. Chapter V, "The Impact of Government on the Evolving Market Structure of the U.S. Wireless Telephony Industry" by Carol C. McDonough, University of Massachusetts Lowell (USA), discusses the impact of government on the market structure of the wireless industry which is regulated by the federal government. Chapter VI, "Shaping Regulation in the Australian Mobile Industry" by Indrit Troshani and Sally Rao Hill, University of Adelaide (Australia), employs qualitative evidence to investigate how regulation and legislation can affect mobile services in the Australian mobile telecommunications industry and proposes an innovative regulatory framework based on a co-regulatory approach.

Section II: Telecommunications Industry, consists of 9 chapters. Chapter VII, "Analyzing the Disruptive Potential in the Telecommunications Industry" by Stefan Hüsig, University of Regensburg (Germany), covers the ex ante analysis of potential disruptive technologies. In this chapter, the theory of disruptive technology, the concept of disruptive potential and a method for applying this concept in a telecommunications planning and technology management context is presented. Chapter VIII, "Convergence of the Internet and Telecommunications" by John B. Meisel and Timothy S. Sullivan, Southern Illinois University Edwardsville (USA), provides a framework to explain the convergence of communications networks and analyzes key issues that confront public policymakers. One key competition issue, termed network neutrality, addresses the concern that the evolving broadband network architecture will enable network providers to favor the provider's services or affiliated services at the expense of independent rivals. Chapter IX, "International Internet Interconnection Service in Asia-Pacific Region" by Moon-Soo Kim, Hankuk University of Foreign Studies (South Korea), focuses on current interconnection settlement arrangement models that disfavor ISPs and end-users in the Asia-Pacific region. After reviewing the Internet market and digital divide in the region, the chapter discusses the main current IIS issues between the Asia-Pacific and Western regions into 3 categories of concern: inequity, anticompetitive practices, and the threat of the "balkanization" of the Internet. Chapter X, "A Techno-Economic Analysis of Mobile Virtual Network Operators" by Dong Hee Shin, Penn State University Berks Campus (USA), examines cross-national data in order to identify possible factors related to the observable patterns of Mobile Virtual Network Operator (MVNO) penetration. The study conducts an economic assessment of market structure and environment for different countries' MVNO penetration. Chapter XI, "Evolving Value Networks and Internationalization of National Telecommunication Companies from Small and Open Economies" by Riku Laanti, Fred McDougall, and Georges Baume, The University of Adelaide (Australia), focuses on the internationalization processes of national telecommunications companies (telcos) from small and open economies (SMOPECs) which have moved from a domestic monopoly to an actor within the global industry. Case examples are used to illustrate the internationalization processes of telcos from SMOPECs within the context of the whole industry value network. Chapter XII, "The Effect of Non-Market Strategies in the Mobile Industry" by Zulima Fernández and Belén Usero Sánchez, Universidad Carlos III de Madrid (Spain), analyzes the role of non-market actions (NMAs) in the competitive position of firms in the European mobile telecommunications industry. The findings of an empirical study confirming the effectiveness of legal actions as competition tools in the European mobile telecommunications industry are discussed. Chapter XIII, "Spatiality and Political Economy of the Global Fiber Optics Industry" by Barney Warf, University of Kansas (USA), reviews the historical development of this communications technology. The chapter discusses the spatial distribution of the world's fiber lines, noting major transatlantic and transpacific markets and newer systems. Chapters XIV and XV "Digital Convergence and Home Network Services in Korea: Part 1 - Recent Progress and Policy Implications" and "Digital Convergence and Home Network Services in Korea: Part 2 – Business Models and Strategic Alliances" by Hyun-Soo Han, Hanyang University (South Korea); Heesang Lee, Sungkyunkwan University (South Korea); Yeong-Wha Sawng, Electronic and Telecommunications Research Institute (South Korea), categorize three classes of typical services and five industrial participants with regard to Korea's home network industry. These chapters investigate business interests of the industrial participants comprising convergent home network services, and explore the causes of conflicts among them. They discuss managerial implications and strategic alliance opportunities of the home network service providers in Korea. Strategic propositions for the success of the home network industry are suggested at the firm level with respect to three main converging home network services.

Section III: Telecommunications Business Management consists of 15 chapters. Chapter XVI, "Mixing and Matching Organizational Network Legitimacy Practices to China's Telecommunication Market" by Brian Low, University of Western Sydney (Australia); Wesley J. Johnston, Georgia State University (USA), presents a case research-based framework to describe an iterative and incremental process to help foreign telecommunications firms manage their network legitimacy. Chapter XVII, "Preliminary Knowledge Management Implementation in the Telco Industry" by Chin Wei Chong, Multimedia University (Malaysia); Siong Choy Chong, Putra International College (Malaysia), aims to create a unified model capturing and generalizing the different arrays of preliminary knowledge management (KM) implementation success factors in the telecommunication industry based on the studies in Malaysia. Chapter XVIII, "Activity-based Costing in the Portuguese Telecommunications Industry" by Maria Major, ISCTE – Business School (Portugal); Trevor Hopper, University of Manchester (UK), examines an implementation of an Activity-based Costing (ABC) system in a specific Portuguese telecommunications firm which provided international telecommunications services. The chapter explores how ABC was adopted and used to support decision-making and pricing strategies. Chapters XIX and XX, "The Role of Organizational Culture to the Management of Telecommunication Companies: I. Background and Motivation" and "The Role of Organizational Culture to the Management of Telecommunication Companies: II. Applications and Case Studies" by Antonios D. Kargas and Dimitris Varoutas, University of Athens (Greece), explain what organizational culture is and analyze its importance for the management. They present the theoretical link between organizational culture and a variety of variables, which affect organizational performance and efficiency. Chapter XXI, "Dynamics of Mobile Service Adoption" by Hannu Verkasalo, Helsinki University of Technology (Finland), discusses the adoption of mobile services. The study explores what drives the intention and actual use of some of the emerging mobile services, thus contributing to both industry and academic readers. Chapter XXII, "When Customer Satisfaction Isn't Good Enough: The Role of Switching Incentives and Barriers Affecting Customer Behavior in Korean Mobile Communications Services" by Moon-Koo Kim, Electronics and Telecommunications Research Institute (South Korea); Myeong-Cheol Park, Information and Communications University (South Korea); Jong-Hyun Park, Electronics and Telecommunications Research Institute (South Korea), explains the relationship between customer satisfaction and customer behavior using switching incentives and switching barriers. Chapter XXIII, "Telecommunication Customer Demand Management" by Jiayin Qi, Beijing University of Posts and Telecommunications (China); Yajing Si, Beijing Wuzi University (China); Jing Tan and Yangming Zhang, Beijing University of Posts and Telecommunications (China), proposes a method that telecom operators can use to estimate their customer demand and respond to their demand automatically. Chapter XXIV, "Telecommunication Customer Detainment Management" by Jiayin Qi, Yuanquan Li, Chen Li, Yingying Zhang, and Jing Tan, Beijing University of Posts and Telecommunications (China), proposes an integrated methodological system of telecommunication customer detainment management which includes the customer churn prediction and customer detainment management. It discusses customer detainment management strategies. Chapter XXV, "How to Plan for an Upgrading Investment in a Data Network?" by Frédéric Morlot and Salah Eddine Elayoubi, Orange

Labs (France), illustrates how a mobile data network operator can plan an upgrading investment to anticipate explosions of the demand, taking into account the expected profit and the customer satisfaction. The chapter then proposes and discusses two methods that help in making the decision: an actualization algorithm and a real option-like strategy. Chapter XXVI, “Telecommunication Investments Analysis: A Multi-Criteria Model” by Georgios Angelou and Anastasios Economides, University of Macedonia (Greece), deals with quantitative and qualitative analysis and integrates real options (ROs) and Analytic Hierarchy Process (AHP) into a common decision analysis framework providing a multi-criteria model for analyzing telecommunication investments in the deregulated business field. Chapter XXVII, “A Business Planning Framework for WiMAX Applications” by Rebecca De Coster, Brunel University (UK), introduces a business planning framework which identifies generic approaches for examining the market potential for WiMAX applications. Chapter XXVIII, “Adoption of VoIP Applications in Public and Private Organizations” by Nicoletta Corrocher, Bocconi University (Italy); Roberto Fontana, University of Pavia (Italy); Claudia Parlanti, Bocconi University (Italy), investigates the determinants of diffusion of VoIP applications in a sample of public and private organizations in Italy. Results suggest that organizations become more likely to adopt as time goes by, and that the decision to adopt is mostly affected by size and availability of financial resources. Chapter XXIX, “Intelligent Networking and Business Process Innovation: A Case Study Analysis of Home Box Office and Dell Computers” by Richard A. Gershon, Western Michigan University (USA), examines the subject of business process innovation which involves creating systems and methods for improving organizational performance. Special attention is given to the topic of intelligent networking which represents the combination of software, technology and electronic pathways that makes business process innovation possible for both large and small organizations alike. Chapter XXX, “Can M-commerce Benefit from Pervasive Computing?” by Stan Kurkovsky, Central Connecticut State University (USA), discusses a number of pervasive computing principles and illustrates how they have been implemented in mobile commerce applications. The chapter also presents new trends in developing context aware m-commerce services that tap into the power of Web 2.0 services and digital communities.

Section IV: Telecommunications Technology Management consists of 12 chapters. Chapter XXXI, “Basics of Telecommunication Management” by Katalin Tarnay, University of Pannonia (Hungary), familiarizes the reader with the main components, functions, and key terms of telecommunications management. Starting at the definition of telecommunications management, it presents the management functions, the managed objects, the management information bases, and the main management protocols in a simple understandable way. Chapter XXXII, “Telecommunication Management Protocols” by Katalin Tarnay, University of Pannonia (Hungary), presents the reader with the fundamentals of communication protocols in telecommunications management. It presents the most widely used protocol, the Simple Network Management Protocol (SNMP), the open system management protocols, and the mobile internet management protocols for Authentication, Authorization and Accounting (AAA). Chapter XXXIII, “Cellular Network Planning- Evolution from 2G to 4G” by Marc St-Hilaire, Carleton University (Canada); Samuel Pierre, École Polytechnique de Montréal (Canada), presents the characteristics and the architecture of each generation of cellular networks (1G, 2G, 3G and 4G). Then, it exposes different planning problems related to each generation followed by a short description of different solutions that have been proposed in the literature. Chapter XXXIV, “IP Address Management: Challenges, Solutions and Future Perspectives” by Faouzi Kamoun, University of Dubai (UAE), discusses the most important challenges in IP Address Management (IPAM) and attempts to address some of the solutions and best practices to tackle these challenges. The chapter underlines the need for organizations to adopt proven

IPAM best practices. Chapter XXXV, “Digital Cable TV Networks: Convergent Technologies, Value-added Services and Business Strategies” Ran Wei, University of South Carolina (USA) and Zizhong Zhao, Communication University of China (China), focuses on digital cable TV networks as a convergent network that provides broadcasting TV and radio, telecommunications services, and IP-based publishing and E-Commerce. This chapter examines the value chain and collaborative opportunities among the participants in the digital cable TV revolution. Chapter XXXVI, “Diffusion and Oscillation of Telecommunications Services: The Case of Web 2.0 Platforms” by Tobias Kollmann, Christoph Stöckmann, and Carsten Schröer, University of Duisburg-Essen, Campus Essen (Germany), focuses on the diffusion characteristics of Web 2.0 Platforms and provides management concepts for competition. The chapter shows that there is always a chance to capture a market or at least to grow against competition in a Web 2.0 setting. Chapter XXXVII, “The diffusion of WiMax Technology: Hurdles and Opportunities” by Phillip Olla, Madonna University (USA), discusses the potential of WiMax technology to deliver personal broadband and fixed Internet capabilities. The chapter applies components from the global diffusion of the Internet framework to analyze the current state of WiMax deployment. Chapter XXXVIII, “VoIP Quality and Security Issues for Consumers and Small Businesses” by David Bell, Pacific Union College (USA); Sunil Hazari, University of West Georgia (USA), reviews issues related to the quality and security of VoIP faced by consumers and small businesses. Chapter XXXIX “Emerging Security Issues in VANETS for E-Business” by S. S. Manvi, REVA Institute of technology and Management (India); M. S. Kakkasageri, Basaveshwar Engineering College (India), presents the emerging security issues in Vehicular Ad hoc Networks (VANETs) for E-Business, along with some of the solutions. Security issues discussed are authentication, trust, confidentiality, non-repudiation and detection of malicious users. Chapter XL, “Telecommunications Network Planning and Operations Management in an Academic Environment: The Case Study of the Aristotle University of Thessaloniki” by Sotirios K. Goudos, Angeliki Z. Agorogianni, and Zaharias D. Zaharis, Aristotle University of Thessaloniki (Greece), presents and discusses network management and planning issues in the Aristotle University of Thessaloniki, Greece. The network expansion and cost reduction case studies are discussed. It also discusses the migration analysis for future upgrades that will fully enable the use of emerging technologies such as Voice over IP (VoIP). Chapter XLI, “Planning of Wireless Community Networks” by Karoly Farkas, University of West Hungary (Hungary); Csaba A. Szabó and Zoltán Horváth, Budapest Univ. Tech&Econ (Hungary), discusses the applications, state-of-the-art technologies, planning methods, and business models for wireless community networks, and provides an integrated presentation of these essential parts with an ongoing digital city project in Hungary. Chapter XLII, “Human Factors for Business Mobile Systems” by David Golightly, University of Nottingham (UK), addresses human factors with particular relevance to developing and deploying business mobile systems. This chapter argues that it is critical to take an approach that puts users at the heart of a new technology or service development. This approach is presented as a three-stage model of context analysis, specification and design, and evaluation

Section V: Telecommunications Technology and Applications consists of 11 chapters. Chapter XLIII, “Towards QoS-Inferred Internet” by Mohamed Boucadair, Pierre Levis, and Pierrick Morand, France Telecom-Orange Labs (France), focuses on two issues: provider-to-provider agreements and enhancements to inter-domain routing protocol to convey QoS-related information. Chapter XLIV, “The Playout Control Management: An Issue for the IP Telephony Service Providers” by Mirko Luca Lobina, Luigi Atzori, and Fabrizio Boi, Universita’ di Cagliari (Italy), proposes playout control management and planning strategies and compares the strategies IP Telephony Service Providers can choose with the aim of saving money and offering a better quality of service. Chapter XLV, “Open APIs and Protocols for Services and Applications in Telecoms" by Nikolaos D. Tselikas, National Technical University of Athens (Greece); Georgia M. Kapitsaki, National Technical University of Athens (Greece); Dimitrios G. Makris, University of Peloponnese (Greece); Iakovos S. Venieris, National Technical University of Athens (Greece), discusses the role of open APIs and protocols for advanced service provisioning. Specifically, the role and the trade-offs in modern telecoms between open APIs and Protocols, that is OSA/Parlay APIs, JAIN APIs and SIP, are discussed. Chapter XLVI, "On the Management Performance of Networked Environments Using Web Services Technologies" by Lisandro Zambenedetti Granville, Ricardo Neisse, Ricardo Lemos Vianna, and Tiago Fioreze, Federal University of Rio Grande do Sul (Brazil), provides a comparison between different strategies of integrating conventionally managed network devices into Web Services-based management processes. Chapter XLVII, "Automated Fault Management in Wireless Networks" by Raquel Barco and Pedro Lázaro, University of Málaga (Spain), proposes a system for automated fault management in the Radio Access Network (RAN). Some methods proposed for auto-diagnosis are also discussed. Chapter XLVIII, "Modeling Intrusion Detection with Self Similar Traffic in Enterprise Networks" by Cajetan M. Akujuobi and Nana K. Ampah, Prairie View A & M University (USA), proposes IDS technique, the novelty of which lies in its efficiency and ability to eliminate most limitations of existing IDSs and IPSs, thereby ensuring high-level network protection. Chapter XLIX, "High-speed Multimedia Networks: Critical Issues and Trends" by Dimitris Kanellopoulos, University of Patras (Greece), presents high-speed networking technologies and standards such as ATM, Fast Ethernet, 10 Gigabit Ethernet, SONET, RPR, PBT, PBB, T-MPLS and Optical Transport Network (OTN). It considers the requirements imposed to high-speed networks by multimedia applications and analyzes crucial issues of high-speed networking such as bandwidth problems, discarding policies and fast broadcast. Chapter L, "Emerging Telecommunications Technologies: Cognitive Radio" by J. Joaquín Escudero-Garzás and Ana García-Armada, University Carlos III de Madrid (Spain), introduces the novel concept of Cognitive Radio for wireless telecommunications. Cognitive radios are a new type of radio devices that include cognition and reconfigurability features. Chapter LI, "Game Theory for PHY layer and MAC sublayer in Wireless Telecommunications" by J. Joaquín Escudero-Garzás and Ana García-Armada, University Carlos III de Madrid (Spain), addresses the role of Game Theory for wireless communications. It introduces Game Theory and its fundamentals, and gives an overview of the applications of Game Theory in the wireless networks for PHY layer and MAC sublayer. Chapter LII, "Agent Based Product Negotiation Models in Mobile Commerce" by S. S. Manvi, REVA Institute of technology and Management (India); L. B. Bhajantri, Basaveshwar Engineering College (India), addresses various issues in m-commerce and presents various agent based product negotiation models in the mobile commerce environment. Chapter LIII, "Role of Telecommunications in Precision Agriculture" by James M. McKinion, USDA-ARS (USA), discusses the various data/information intensive precision agriculture applications which use wireless local area networking and Internet access, and the next generation technologies which can immensely propel precision agriculture to widespread use in agriculture.

Section VI: Telecommunications Network Design consists of seven chapters. Chapter LIV, "Network Planning and Dimensioning for Broadband Access to the Internet Regarding Quality of Service Demands" by Franz Hartleb, Gerhard Hasslinger, and Sebastian Kempken, Universität Duisburg-Essen (Germany), focuses on planning and traffic engineering for link bandwidth and buffers as main resources in communication platforms based on measurement and statistical properties of traffic growth and variability. Chapter LV, "Overlay Networks: New Techniques for Global Service and Network Provisioning" by Gerhard Hasslinger, Deutsche-Telekom (Germany), discusses overlay networks within the standardization of Internet Protocols (IP), where the main focus is on evolving techniques on higher layers. Chapter LVI, "Optimising P2P Networks by Adaptive Overlay Construction" by Nick Antonopoulos and James Salter, University of Surrey (UK), proposes ROME, a layer which runs on top of the Chord DHT to provide control over network size through the monitoring of node workload and suggests the use of processes to control the addition or removal of nodes from the network. Chapter LVII, "Designing a New Service Overlay on a Carrier Network Using the Efficient Frontier" by Susan J. Chinburg, Rogers State University (USA); George Scheets, Oklahoma State University (USA), expands the optimal network design process by applying a process borrowed from Finance and Operations Research/Management Science literature known as the Efficient Frontier or production frontier analysis. Chapter LVIII, "Combining Small and Large Scale Roaming Parameters to Optimize PCS Networks Design" by Mohamed Zaki and Salah Ramadan, Azhar University (Egypt), introduces a model that unites both an Enhanced Profile-Based Strategy (EPBS) for small-scale roaming and a Caching Two-Level Forwarding Pointer (C2LFP) strategy. The idea behind this model is applying those two location management strategies to specify the physical parameters of PCS networks from mobility management's point of view. Chapter LIX, "Secure Resource Optimization in Distributed Service Computing" by Kaiqi Xiong, Texas A&M University Commerce (USA); Harry Perros, North Carolina State University (USA), presents a framework for optimizing a set of computer resources used by a service provider to host enterprise applications subject to a service level agreement (SLA). Chapter LX, "Trends, Optimization and Management of Optical Networks" by Jadranka Skorin-Kapov, Stony Brook University (USA); Darko Skorin-Kapov, Adelphi University (USA), provides a review of optical networking, its current issues, trends, and ways to efficiently design and manage a wide range of optical networks.

The *Handbook of Research on Telecommunications Planning and Management for Business* is an excellent collection of the latest research and practices associated with emerging telecommunications policies, management, technologies, and applications. This book is the first comprehensive book that presents aspects from the government, industry, managerial, and technical sides of the telecommunications. As leading experts in the telecommunications area, the contributors did an excellent job of providing our readers with extensive coverage of the most important research topics, concepts, business practices, technologies, and trends in the telecommunications industry. The projected audience includes policymakers, telecommunications employees, managers, researchers, professors, and undergraduate/graduate students in various telecommunication management programs. I expect this book to shed new insights for researchers, educators, and practitioners to better understand the important issues and future trends of telecommunications research and technologies.

REFERENCES

Bauer, J.M. and Bohlin, E. (2008). From Static to Dynamic Regulation - Recent Developments in US Telecommunications Policy. *Intereconomics*, 43(1), pp. 38-50.

Tallberg, M., Hämmäinen, H., Töyli, J., Kamppari, S., & Kivi, A. (2007). Impacts of handset bundling on mobile data usage: the case of Finland. *Telecommunications Policy, 31*(10/11), 648-659.

Ubacht, J. (2004). Regulatory practice in telecommunications markets: how to grasp the 'real world of public action'? *Communications & Strategies*, (54, 2nd Quarter), 219-242.

Acknowledgment

I would like to acknowledge the help of the eighty-four expert reviewers involved in the double-blind review process for sixty chapters of the handbook, without whose support the book could not have been satisfactorily completed.

I sincerely thank Ms. Kristin M. Klinger, Julia Mosemann, Heather A. Probst, Jan Travers, Mehdi Khosrow-Pour, and other members of the IGI Global, whose contributions throughout the whole process from the inception of the initial idea to the final publication have been invaluable.

In closing, I would like to express my gratitude to the 111 authors for their invaluable contributions and collaboration.

In Lee, Ph.D.
Western Illinois University
Macomb, IL, USA
August 2008

Section I
Telecommunications Policy

Chapter I
Static and Dynamic Efficiency in the European Telecommunications Market:
The Role of Regulation on the Incentives to Invest and the Ladder of Investment

Walter Distaso
Imperial College London, UK

Paolo Lupi
AGCOM, Italy

Fabio M. Manenti
Università di Padova, Italy

ABSTRACT

In this chapter the authors provide evidence of the effectiveness of European National Regulatory Authorities in applying the basic principles of the ladder of investment. The analysis discusses and compares the regulatory approach adopted in 12 European Countries from January 2005 to July 2007. Results are not easy to interpret, given the peculiarities of the different national markets. Nevertheless, they can conclude that the policies adopted by National Regulatory Authorities are broadly consistent with the ladder of investment theory.

1. INTRODUCTION TO THE THEORY OF INVESTMENT LADDER

With the advent of next generation networks, the telecommunication sector is undergoing a deep transformation: from an industry characterized by a "quasi-monopolistic" infrastructure to an industry where alternative networks compete with the incumbent (and with each other). It is widely accepted that these changes are actually challenging the regulatory approach towards telecommunications: while in the near past regulators' efforts were aimed at introducing more competition into quasi-monopolistic markets, nowadays regulation also tries to create favourable conditions for the necessary capital investments needed to sustain and innovate the network infrastructure (see, among others, Bauer and Bohlin, 2008).

This transformation also exacerbates the traditional trade-off between static and dynamic efficiency. Static efficiency occurs when marginal production costs are minimized (production efficiency) or when the price consumers pay in exchange of a good or service equals the production cost (allocative efficiency); on the other hand, dynamic efficiency relates to demand creation and innovation. While in competitive markets firms price at cost thus achieving static efficiency, competition may actually be the source of dynamic inefficiency since it reduces the incentives to innovate by preventing innovators from recovering the fixed costs of their investment in new technologies.

Usually, industry regulators face this trade-off whenever they attempt to maximize static social welfare through price regulation, while providing firms with sufficient incentives to innovate (see, e.g., Guthrie, 2006).

All these issues are particularly important in telecommunications, probably the most dynamic industry subject to sector specific regulation (for a theoretical and empirical analysis of the impact of regulation on the dynamic efficiency in telecommunications, see Prieger and Heil, 2008). Here, the tension between static and dynamic efficiency is so crucial that it is explicitly mentioned in the new European regulatory framework for electronic communications services. There, it is clearly stated that the scope of regulation is to "promote competition in the provision of electronic communications networks, electronic communications services and associated facilities" and to "encourage efficient investment in infrastructures and to promote innovation" (Framework Directive, Article 8.2.).

Regulation may influence innovation in telecommunications either through prices or by impacting the dynamics of market entry (Bourreau and Dogan, 2001). In telecommunications, both retail and wholesale (interconnection) charges are heavily regulated. This has an impact on operators' profits and, consequently, on their incentives to invest in innovative activities. At the same time, regulation may alter the terms of entry (which are often asymmetric, in the sense that entrant operators are often not subject to regulatory impositions) and, again, this is likely to have an impact on the degree of innovation undertaken by both the incumbents and entrant operators.

In order to reconcile static with dynamic efficiency, European countries have implemented a set of common regulations based on the key regulatory instrument of local loop unbundling (LLU). LLU is the regulatory process whereby incumbent operators lease, wholly or in part, at a regulated price the local segment of their telecommunications network (usually pairs of copper wire) to competitors to let them provide voice and broadband services on the retail market.

Regulatory provisions aimed at implementing LLU have been introduced in all member states since 2001. The regulatory framework adopted by the national regulatory authorities of all EU15 member states mandates incumbent operators the provision of a (more or less) wide portfolio of wholesale services that alternative operators aggregate according to their needs in order to provide voice and broadband services to their customers.

More specifically, local loop unbundling usually comes in three forms: bitstream access, shared access and full local loop unbundling. Bitstream access occurs when the incumbent installs a high speed access link to the customers premises (e.g., by installing ADSL equipment in the local access network) and then makes it available to third parties. With bitstream access, entrants do not have control over the physical line nor are they allowed to add other equipment; therefore, entrants are restricted to resale the services designated by the incumbent operator (usually broadband internet access). Bitstream is a form of wholesale access usually chosen by less infrastructured alternative operators, whose proprietary network interconnects with the incumbent's network through a limited number of higher-level points of interconnection.

Alternatively, entrants may gain access via either shared access or full LLU. With shared access, the incumbent remains in control of the copper line and can still provide (voice) services to consumers, since entrants only lease part of the copper pair spectrum (the high-frequency, non voice spectrum). With fully unbundled access, the alternative operator obtains full control of the copper line and this allows him to offer both voice and data services. Shared access and, particularly, LLU are technically and economically viable only to operators with a sufficiently widespread network, able to reach most of the local exchanges of the incumbent. For this reason, they are usually the access option preferred by infrastructured operators.

European countries have implemented LLU in order to favour entry by alternative operators and to reduce incumbents' market power. Nevertheless, in line with the aforementioned tension between static and dynamic efficiency, many feared that wholesale access could undermine firms' incentives to invest and to innovate. The argument goes as follows: LLU erodes incumbents' profits and, consequently, their incentives to invest in new technologies and services. On top of that, by allowing entry without the need of a proprietary infrastructure, LLU would not provide incentives to invest by the entrant firms either.

To solve this potential standstill, Martin Cave and Ingo Vogelsang have proposed a new approach to LLU regulation, known as the "ladder of investment" theory (see Cave and Vogelsang, 2003 and Cave, 2006). This theory is based on the idea that in telecommunications, initially essential facilities may become less essential over time; this argument is not new in the literature (see, among others, Bourreau and Dogan, 2006),[1] but the authors have reconsidered it to provide a novel approach to telecoms regulation.

The mechanism underlying the ladder of investment is extremely simple. At the beginning, the regulator should encourage access to wholesale markets by setting very low access prices for the network elements that are too expensive for new entrants to replicate. As soon as new entrants consolidate their market position and have gained brand recognition and customer base, regulatory authorities should increase access prices, starting from the network elements that are easier to duplicate (i.e. bitstream access). The price increase of these network elements should encourage new entrants to invest on these elements, in order to migrate towards higher levels of the investment scale in infrastructure, i.e. to move up the "ladder of investment". Ultimately, as soon as entrant operators have gained sufficient revenues and technological know-how, they can finally climb the last step of the ladder, i.e. they can invest in their own access infrastructure.

In order for the investment ladder to work effectively and to induce operators to climb the ladder, national regulatory authorities should adopt what has been defined as an increased micro-management of access products (Cave, 2006), a carefully designed dynamic access regime.

The ladder of investment theory has been extremely appealing to industry regulators and the European Commission since it offers a possible solution to the trade-off between static and dynamic efficiency; in fact, due to mandated LLU, entry and competition are guaranteed in the short run while operators are given the right incentives to invest (namely, to climb the ladder) through a dynamically appropriate design of access terms.

The investment ladder has also some opponents, expressing doubts with regards to, at least, two aspects of the theory (see Oldale and Padilla, 2004). They argue that, on one side, it is doubtful that regulators are able to implement the required micro-management of access terms. Furthermore, it is not clear why LLU-oriented *service-based* competition should necessarily evolve into a robust *facility-based* competition.

Finally, from an empirical point of view, the real efficacy of the theory of the ladder of investment has never been established.

2. EMPIRICAL ANALYSIS

The aim of this section is to investigate empirically the ladder of investment theory. We start by surveying the most recent literature dealing with regulation and investment in telecommunications. As we will discuss briefly, the various empirical studies often reach different conclusions on the existence and the strength of the trade-off between static and dynamic efficiency in telecommunications. Furthermore, they often present various limitations and shortcomings that call for further investigations. In the second part of this section, we use a novel and detailed dataset on access lines and prices to check the validity of the investment ladder in various European countries.

2.1. Survey of Recent Studies

The relationship between regulation and investments in telecommunications has been recently analyzed in a series of papers. Friederiszick et al. (2008) use firm-level data on tangible fixed assets as a measure of investment in telecommunications. Their sample is made of 180 firms in 25 European countries and the study covers the period 1996-2007. The use of a detailed data-set allows the authors to distinguish the effect of regulation between entrant and incumbent firms. Regulation is proxied by the Plaut's regulatory index, which is explicitly constructed to account for the influence that a regulatory intervention may have on firms' investment activities. The authors find no evidence of the investment ladder. More specifically, they find that entry regulation discourages infrastructure investment by entrant firms in fixed-line telecommunications and does not significantly affect incumbents' investment behavior. In other words, this study confirms the existence of the trade-off between static and dynamic efficiency highlighted in the theoretical literature. No evidence of the investment ladder theory is also found by two earlier papers, albeit not based on a systematic econometric analysis: Hausman and Sidak (2005), which focuses on the US, the UK, New Zealand, Canada and Germany, and Hazlett and Bazelon (2005), which concentrates on the US.

A rather different result is reached by Jones and Salsas (2006), who present the outcome of a study commissioned by the European Commission's DG Information Society aimed at identifying the main drivers of e-communications investments in the EU. As in the previous case, investment data are taken from annual reports of over 200 firms for the period 2001-04. For each country, the regulatory activity is proxied using the OECD regulatory index, which measures regulatory performance in terms of *i)* the degree of free entry into the market, *ii)* the extent of government ownership of the major operators and *iii)* market structure based on market shares. The index comes out to be significant, suggesting that those countries with lower performances in terms of regulatory activity (namely, less competition/entry and more government ownership of incumbent firms) are those that have less investment.

On similar grounds, Cadman and Dineen (2006) investigate the relationship between regulation and investments in 9 European countries over 3 years. To measure the effectiveness of regulation in telecommunications, the authors use the ECTA scorecard, a composite index introduced by the European Competitive Telecommunications Association (ECTA). ECTA aggregated the responses submitted by national regulatory authorities and companies to a detailed questionnaire covering institutional framework and various market conditions (both competitive and regulatory conditions). The study finds a statistically significant relationship between regulatory effectiveness and investments, suggesting that in those countries where regulation supports market entry, a higher level of investments in telecommunications occurs.

These studies reach different conclusions: in particular, the latter ones seem to contrast the traditional view, discussed previously, based on the existence of a trade-off between static efficiency and investment in fixed telecommunications. Nevertheless, all these results need to be taken cautiously for at least two important reasons: on the one side, these papers use investment data taken from companies' annual reports. As discussed in OECD (2007): "The quality of data on investment by new entrants in telecommunication markets is poor. Definitional problems also exist in that the players in communication markets are no longer the traditional alternative operators providing voice services but a range of service providers including Internet Service Providers and cable television service providers who, through cable modems, provide VoIP services and broadband access". This implies that investment data taken from annual reports can be a very poor indicator of operators' real amount of investment in new technologies.

Furthermore, they all use aggregated indices of regulatory effectiveness: these indices summarize various policy instruments and are often constructed in a very *ad hoc* way. As a result, it is rather questionable whether they effectively reflect the actual intensity of regulation in a given country (see Weeks and Williamson, 2006 for a series of arguments against the ECTA scorecards).

A different approach is followed in Waverman et al. (2007), in a study conducted on behalf of the association of the European Telecommunications Network Operators (ETNO). The authors test the impact of specific aspects of access regulation embodied in the price of LLU on investment in alternative access platforms. Unlike previous studies, this econometric model does not rely on indices or scorecards extrapolated in some way from market data and it does not use data on investments obtained from companies' annual reports. Regulation is measured directly using data on LLU access prices while investments are proxied using growth rates in unbundled lines relative to *facility-based* lines, i.e. lines based on alternative infrastructures. The analysis goes from 2002 to 2006 and covers 27 European countries. The authors find evidence that lower local loop prices cause strong substitution from broadband offers over alternative platforms towards LLU-based broadband offerings and this substitution ultimately results in substantially lower investments in alternative access platforms. Interestingly, these findings move in a direction which is not compatible with that envisaged by the ladder of investment; although the authors are not completely critical of the concept of the ladder of investment, they make the reader aware of the risk associated with an improper implementation of this regulatory approach in Europe: as they suggest "the presumption of an enduring access bottleneck may tempt regulators into continuing to promote competition at lower rungs of the ladder (by offering access to incumbent networks on attractive terms) and thus result in a technologically stagnant marketplace."

2.2. Results

Similarly to Waverman et al. (2007), we test the ladder of investment theory by looking at the link between the prices of wholesale access services and the relative growth rates of the three alternative inputs that can be used by new entrants to provide access and broadband services to end users: bitstream services, LLU services and their own network. In other words, we overcome the difficulties associated with obtaining disaggregated data on investment in telecommunications infrastructures carried out by new entrants in each country, by proxying the level of this investment with the shares of the inputs used by new entrants to provide retail services.

We track for each country the evolution of bitstream lines, unbundling lines and proprietary access infrastructures. If the ladder of investment is consistently applied, we might expect to see a migration from access services, like bitstream and resale, that require a low level of infrastructures – and therefore investment – to services, like shared access and LLU, that require an higher level of infrastructures and investment. We should also observe the share of LLU services to increase with respect to the share of bitstream services. Finally, and more importantly, we should expect an increase in the share of proprietary infrastructures, to reflect a move towards that kind of infrastructure.

As discussed earlier, the Ladder of Investment (LoI) theory implies that National Regulatory Authorities (NRAs) follow a carefully designed dynamic access regime which, initially, requires to regulate the price of bitstream services at a very low level in order to promote service based entry. Once entry has occurred, NRAs should stimulate infrastructure investment (i.e. to climb the ladder) by *a)* increasing the price of bitstream services and *b)* decreasing the price of LLU services, a more infrastructured form of access. If the regulatory authorities follow a LoI regulatory approach, we should expect the ratio between LLU and bitstream access to decrease through time. Finally, in order to incentivize firms to develop proprietary networks, NRAs should progressively increase the price of LLU services.

In order to graphically explore the relationship between investments and wholesale access prices, we show, for each country, two diagrams. The first is a so called "ternary diagram" plotting the evolution over time of the shares of the three inputs that, at an aggregate level, are used by new entrants to provide retail access and broadband services. The second is a plot of the ratio between the percentage changes (since 2005) in the regulated LLU price and the price of bitstream access.

We use ternary diagrams because they allow to map in a two dimensional space the relative share of three variables – the three inputs – that sum up to a constant (one in our case). The share of each input is represented by the position of a point on an equilateral triangle.[2] Each vertex of the triangle represents a distribution of inputs that puts all the weight on a single input, for instance the right vertex of Figure 1 represents a situation where firms use only ULL in order to provide retail access services. Points on the edges of the triangle represent situations where firms use only two of the inputs (the inputs that are on the two vertices connected by the edge), for instance points on the lower edge of the triangle in Figure 1 represent a situation where firms use a combination of bitstream and ULL services. Finally, points in the interior of the triangle represent a situation where firms use a combination of all three inputs. Of course, the closer one point to one vertex of the triangle, the higher the share of that input.

In order to test whether NRAs use their regulatory instruments (the prices of mandated access services) according to the principles and the objectives of the ladder of investment, we have plotted the ratio between the percentage change in the price of an unbundled loop from the base year (2005) and the price of a retail ADSL line.[3] We have proxied the price of a bitstream line with the price of the cheapest DSL offer of the incumbent operator because, given the complexities and peculiarities of the bitstream reference offers of European incumbent operators, it is almost impossible to derive for each country and for each time period a single representative price. We think that the price of a retail ADSL offer of the incumbent is a good proxy of the price of a bitstream line for several reasons. Firstly, during our period of investigation, almost all European NRAs were mandating bitstream prices based on the retail minus methodology that links the price of bitstream line to the price of a retail line. In addition, the value to be subtracted from the retail price, in order to get the wholesale bitstream price, is usually constant over time and does not differ significantly across countries. Lastly, in our case, we are more interested in variations in prices than in their absolute levels: variations in retail prices are immediately transmitted to bitstream prices through the retail minus methodology.

If the LoI is in place we should expect that the dynamic path of inputs moves from a point close to the left vertex (area 1 in Figure 1) to a point close to the upper vertex (area 3 in Figure 1), passing through the right vertex (area 2 in Figure 1), thus indicating that entrants are effectively substituting resale and bitstream access with more infrastructured forms of access, that is they are "climbing the ladder". On the other hand, if regulation is consistent with the investment ladder, NRAs should trigger the relative variation of the prices in order to incentivize new entrants to climb the ladder. Hence, in the price diagram we should observe a decrease in the ratio between LLU price variations and bitstream price variations.

For our visual analysis of the investment ladder, we have combined data coming from different sources. All data on the number of resale, bitstream, fully unbundled and shared access lines (as well as the number of DSL, CATV, broadband upgraded CATV, fibre optics and satellite lines) were taken from the "Broadband Access in the EU" reports that are published every six months by the Communication Committee of the European Commission. All data on the prices of fully unbundled loops were taken from the annual reports on the implementation of the telecommunications regulatory package of the European Commission.[4] Data on retail prices for ADSL services of incumbent operators were taken from the Communication Outlook and the "Broadband Portal" of the OECD.[5] For each country and each period we have chosen the price of the cheapest ADSL offer, with a download speed not lower than 640 Kbit/s, of the incumbent telephone operator.

The dataset contains information for the following countries: Austria (AT), Belgium (BE), Denmark (DK), France (FR), Finland (FI), Germany (DE), Italy (IT), Netherlands (NL), Portugal (PT), Spain (ES), Sweden (SE) and United Kingdom (UK). We have half yearly observations from January 2005 to July 2007, the latest available figure.

In order to better interpret and put into context the graphical information contained in the ternary and price plots, in Table 1 we also provide a measure of competition (denoted by Inc., the share of the incumbent operator in the broadband market), and a measure of penetration of the service (denoted by Pen., the ratio between the number of active broadband lines and the population).

Our dataset starts in January 2005, when in most central European countries broadband services were already available since a few years; clearly at that time, the various countries were at different stages of the dynamic process implied by the ladder of investment. In other words, the starting point indicating the share of bitstream services, LLU services and own network used to provide access and broadband services by new entrants in January 2005 differed across countries. The fact that countries were at different stages (or

Figure 1. How to interpret ternary diagrams

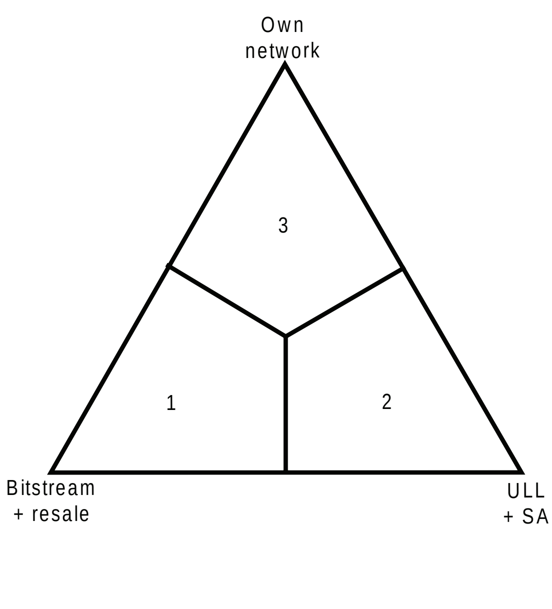
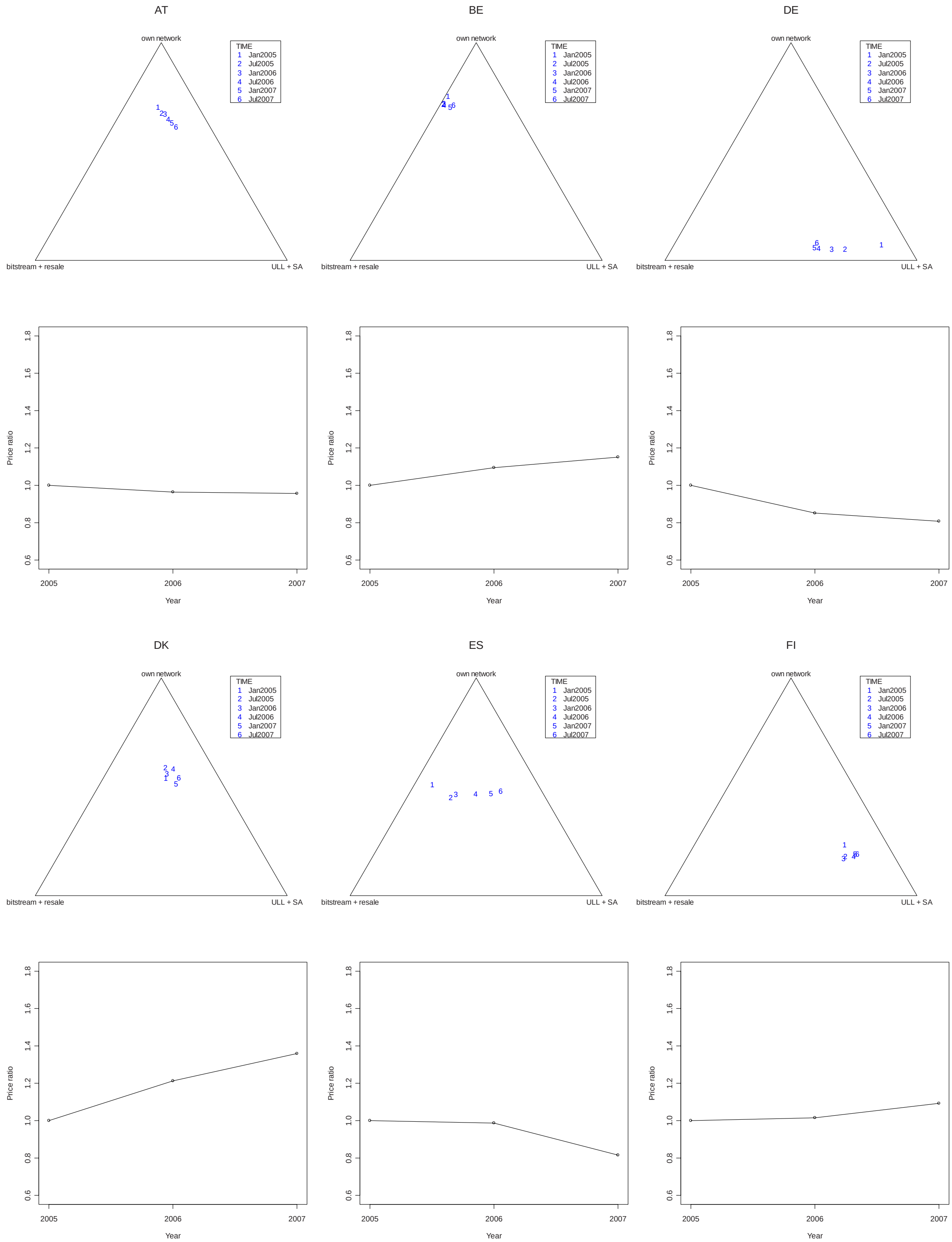


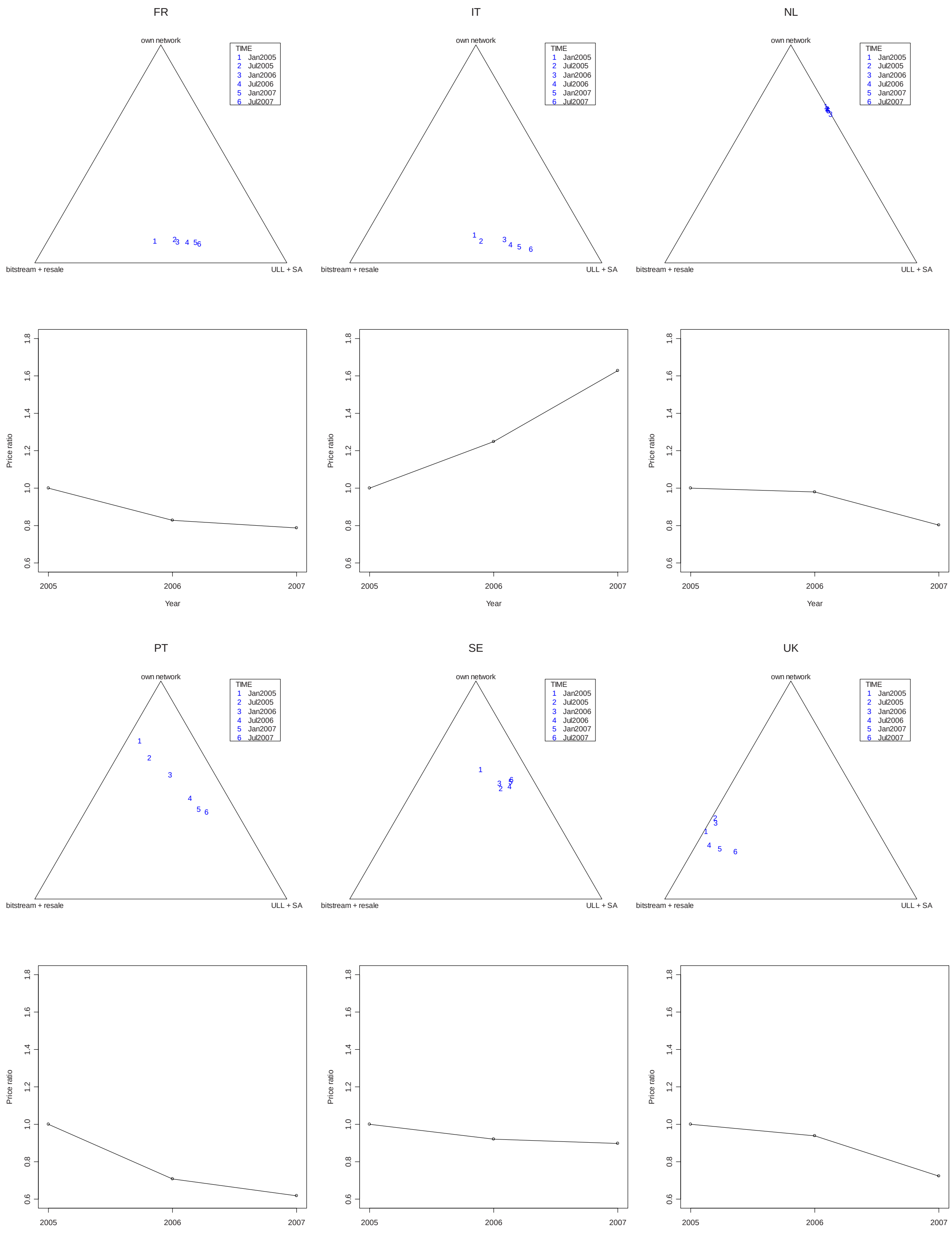
Figure 2. Ternary and price plots



continued on following page

8

Figure 2. continued



different rungs) of the ladder of infrastructure puts us in a position to check whether the principles of the ladder of investment were implemented consistenly. Therefore, we first need to identify for each country where entrants were located on the investment ladder at the beginning of 2005.

This distinction is crucial since, following the arguments proposed in Cave (2006), in order to make the ladder of investment operational, a NRA needs first to assess where on the ladder all firms are located to avoid a fundamental danger – namely that the ladder approach will slow down rather than speed up competitive investment.

Following the input segmentation given previously, we can identify three types of countries: *i*) countries that in January 2005 fall in region 1 of Figure 1 (Type I countries), characterized by high adoption rates of resale and bistream services, and therefore by a low level of investments in LLU and own infrastructures. *ii*) Countries that in the same period fall in region 2 of Figure 1 (Type II countries), characterised by low levels of proprietary infrastructures but by a relatively large share of LLU and shared access services (at least when compared to Type I countries). *iii*) Countries that at the beginning of the sample fall in region 3 of Figure 1 (Type III countries), characterised by the highest level of proprietary infrastructure.

Once countries have been classified as type I, II or III, it is then possible to look at the evolution of the shares of the three inputs in the period January 2005-July 2007 and then to check whether in each country the regulatory approach implemented by the national regulator, represented by the price plots, is consistent with the dynamic access regime implied by the ladder of investment theory.

From a simple inspection of the ternary diagrams it follows that Spain, France and Italy can be clasiffied as Type I countries. Germany and Finland, being characterized by a slightly higher level of infrastructured entrants, can be classified as Type II countries, while all remaining countries, having a rather large share of fully infrasructured entrants, can be classified as Type III countries.

Type I Countries (Spain, France and Italy)

These countries are characterized by a very low level of access infrastructure which can be mainly ascribed to the absence (or the limited coverage)

Table 1. Broadband penetration and competition in selected EU countries (%)

Country	Jan 2005		Jul 2005		Jan 2006		Jul 2006		Jan 2007		Jul 2007	
	Pen.	Inc.	Pen.	Inc.	Pen.	Inc.	Pen.	Inc.	Pen.	Inc.	Pen.	Inc.
AT	10	36.1	12	37.2	14.3	39.7	16.0	39.7	17.4	39.7	18.4	39.2
BE	16	49.6	17	48.8	19.3	48.3	20.8	48.4	22.4	47.8	23.9	46.8
DE	8	80.4	10	71.6	12.8	60.1	15.3	51.2	18.1	48.8	21.1	46.4
DK	18	58.2	22	55.4	24.7	58.7	29.7	59.7	31.4	63.2	37.2	63.7
ES	8	52.3	10	51.5	11.7	53.3	13.4	55.4	15.2	55.9	16.8	55.8
FI	15	72.3	19	66.8	22.4	66.4	25.0	66.5	27.1	69.7	28.8	69.1
FR	11	44.7	14	44.0	16.4	44.8	18.7	46.5	20.4	46.2	22.3	46.8
IT	8	73.1	10	70.8	11.8	71.7	13.2	67.9	14.5	66.6	15.9	64.8
NL	19	46.1	22	44.5	25.2	44.6	29.1	43.7	31.8	44.8	33.1	44.4
PT	8	80.7	10	79.1	11.6	76.3	12.9	72.1	13.9	71.1	14.8	70.1
SE	15	38.5	17	39.1	20.7	38.4	23.0	39.6	25.9	39.3	28.3	38.9
UK	10	25.1	14	24.8	16.5	24.8	19.4	24.3	21.7	25.2	23.8	25.7

of cable TV networks. It is worth noting that Spain, France and Italy are, among all European countries, the ones where bitstream services have been available for a long time (since 1999). This is most likely due to the fact that there was no alternative platform to provide broadband services and therefore mandating access on the network of the incumbent was the only way to start a competitive dynamics in the country.

In this scenario, the ladder of investment theory suggests to follow the dynamic regulatory access regime described previously, envisaged in a decreasing pattern of the ratio between the early price variation of an unbundled local loop and the price variation of a bitstream line.

Price plots for France and, partially, Spain reveal that the national regulators have followed access regimes which are reasonably consistent with the investment ladder. The ternary plots seem to confirm the ladder of investment, since both countries have experienced a significant move towards a more infrastructured access: although in both countries the share of access through own network remains substantially constant over time, there has been a clear increase in the share of shared access and LLU.

The Italian experience looks a little bit more controversial. In fact, despite a significant increase in the number of shared access and unbundled lines, as in Spain and France, the price plot reveals an access regime which goes in the opposite direction to the one suggested by a correct application of the ladder of investment. In this case, the more pronounced reduction of bistream prices relative to unbundling prices can be ascribed, at least to a certain degree, to the very high starting levels of bitstream prices.

It may be worth noting that among Type I countries (see Table 1), France is the one that shows the highest level of competition and diffusion of broadband services. This result confirms the validity of the principles of the LoI, though it could be due, at least partially, to the more favourable initial position of France. In any case, in 2005-2007 Italy shows the highest increase in competition, with a rise in new entrants' market share of about 8.3%.

Type II Countries (Germany and Finland)

In these countries, in January 2005 new entrants provided broadband access mainly through shared access and local loop unbundling, and very few access lines were offered through proprietary infrastructure.

Despite having a similar wholesale structure at the beginning of 2005, Germany and Finland followed a completely different pattern in terms of both the evolution of access inputs and of policy adopted by the NRAs. Among the two countries, Germany is clearly the one that shows the most peculiar pattern. In fact (see the second and the third column of Table 1), Germany started from a market configuration in January 2005 characterised by a very low level of competition (the market share of the incumbent on the retail market was 80.4%) and by a very low level of penetration of broadband services. This situation is typical of countries where only some of the access products are initially available (at that time there was no mandated bitstream in Germany and a very low amount of resale lines), as market entry requires more initial investment by new entrants. In this context broadband penetration is usually very low, since the incumbent, not being challenged by entrants, inevitably tends to reduce its investment.

We can therefore understand why the German NRA, worried by the disappointing state of the market, has descended the ladder by mandating bitstream services and ensuring the provision of higher levels of resale services. In this way Germany has sensibly increased broadband penetration and competition even if this latter form of competition, being mostly based on resale services, is by many considered to be less effective than competition based on other forms of inputs.

Finland has followed a pattern that is compatible with the principles of the LoI. Looking at the ternary plots, the distribution among inputs has mainly remained constant over the 2005-07 period. This is perfectly consistent with the very limited relative variation of bitstream and unbundling prices. The relative stability of Finland both in terms of prices and inputs is confirmed by the stability of competitiveness and penetration: among the countries in our dataset, Finland's ranking in those two measures remained almost unchanged between 2005 and 2007.

Type III Countries (Austria, Belgium, Denmark, the Netherlands, Portugal, Sweden and UK)

The majority of countries in our dataset are Type III countries, where in January 2005 entrants were endowed with a significant amount of proprietary infrastructure. We can further classify these countries in two groups according to the dynamic patterns that emerge: countries that show no or limited variation in the distribution of inputs, like Belgium, Denmark and Netherlands, and countries that show a sensible variation in the distribution of inputs, like Austria, Portugal, Sweden and UK. The policies adopted by NRAs of both groups are compatible with the ladder of investment.

Countries that belong to the first group are those that in January 2005 (see Table 1) showed the highest level of penetration of broadband services and where coverage of alternative networks, mainly cable TV, was the highest. In these countries, we could say that the mere existence of alternative infrastructures leads (more or less) automatically to competition and in fact Regulatory Agencies (see the price plots), have maintained access prices relatively unchanged.

Countries that belong to the second group started in January 2005 from a lower level of penetration of broadband services. In these countries alternative cable networks had limited regional coverage and influence (at least with respect to first group countries). As a consequence, also alternative operators' ability to compete was limited. Therefore Regulators (see the downward sloping price plots), have used the access price instrument in a way compatible with the principles of the LoI. Recognising and acknowledging the difficulties in further expanding the penetration of services based on alternative infrastructures, they have favoured the development of the DSL part of the market – and the transition from bitstream to unbundling – that, in turn, stimulates competition with the cable segment of the market.

We can therefore conclude that for these countries inter-platform competition is a result of the dynamics of infra-modal competition in the DSL segment of the market and not the cause of a competitive broadband market.

3. CONCLUSION

In this paper/chapter we provide some evidence on the effectiveness of European NRAs in applying the basic principles of the ladder of investment. The results we get from our graphical analysis are not easy to interpret given the peculiarities of the different national markets, but we can conclude that the policies adopted by National Regulatory Agencies are in many, if not most, cases compatible with the ladder of investment.

We are aware that having limited our analysis to the effects on the distribution of inputs induced by National Regulatory agencies' price policies we have not fully covered the extent of the LoI principles. In fact, according to the LoI, Regulators have to support the upward move on the ladder either through dynamic pricing of the access wholesale services or through sunset clauses stating in advance that regulation will be removed and implying that new entrants should not establish themselves forever on a particular rung. However, at this stage, while most Regulators are still erecting the ladder, there are only a few regulators that are, at least on a geographic basis, removing bitstream regulations, and it is too early to anticipate when and in which way regulators will introduce the full set of LoI policies.

REFERENCES

Bauer, J. M., & Bohlin, E. (2008). From Static to Dynamic Regulation - Recent Developments in US Telecommunications Policy. *Intereconomics, 43*(1), 38-50.

Bourreau, M., & Dogan, P. (2001). Innovation and Regulation in the Telecommunications Industry. *Telecommunications Policy*, *25*(3), 167-184.

Bourreau, M., & Dogan, P. (2006). Build or Buy Strategies in the Local Loop. *American Economic Review, Papers and Proceedings* , *96*(2), 72-76.

Cadman, R., & Dineen, C. (2006). *European Telecom's Lost Investment: An analysis of the ECTA Scorecard*. Strategy and Policy Consultants Network Ltd (SPC Network).

Cave, M. (2006). Encouraging Infrastructure Competition through the Ladder of Investment. *Telecommunications Policy*, *30*(3-4), 223-237.

Cave, M., & Vogelsang, I. (2003). How Access Pricing and Entry Interact. *Telecommunications Policy*, *27*, 717-727.

Friederiszick, H., Grajek, M., & Röller, L. H. (2008). *Analyzing the Relationship between Regulation and Investment in the Telecom Sector*. Paper presented at International Industrial Organization Conference, Marymount University, Arlington, VA.

Guthrie, G. (2006). Regulating Infrastructure: The Impact of Risk and Investment. *Journal of Economic Literature*, *XLIV*, 925-972.

Jones, S., & Salsas, P. (2006). E-Communications: Investment and the Regulatory Framework. *Communications & Strategies*, *64*, 29-39.

Hausman, J., & Sidak, G. (2005). Did Mandatory Unbundling Achieve its Purpose? Empirical Evidence From Five Countries. *Journal of Competition Law and Economics*, *1*, 173–245.

Hazlett, T. W., & Bazelon, C. (2005). *Regulated Unbundling of Telecommunications Networks: A Stepping Stone to Facilities-Based Competition?* Paper presented to the Telecommunications Policy Research Conference.

OECD (2007). The Influence of Market Developments and Policies on Telecommunication Investment. *DSTI/ICCP/CISP 6*.

Oldale, A., & Padilla, A. J. (2004). *From State Monopoly to the "Investment Ladder": Competition Policy and the NRF*. In *Swedish Competition Authority, The Pros and Cons of Antitrust in Deregulated Markets*, 51-77.

Prieger, J. E., & Heil, D. (2008). Is regulation a roadblock on the information highway? In I. Lee (Ed.), *Handbook of Research on Telecommunications Planning and Management for Business.* Hershey, PA: IGI Global (pp. 15-32).

Waverman, et. al (2007). Access Regulation and Infrastructure Investment in the Telecommunications Sector: An Empirical Investigation. London Economics Consulting Group with the Support of ETNO, *mimeo*. September 2007.

Weeks, M., & Williamson, B. (2006). *A Sound Basis for Evidence Based Policy? A Critique of the ECTA Regulatory Scorecard and SPC Network Papers on Investment and Broadband*. Paper prepared for ETNO by Indepen.

KEY TERMS

Bitstream Access: It occurs when the incumbent installs a high speed access link to the customer's premises (e.g. by installing ADSL equipment in the local access network) and then makes it available to third parties. With bitstream access, entrants do not have control over the physical line nor are they allowed to add other equipment.

Dynamic Efficiency: It occurs when firms have an incentive to invest in innovation (either product or process innovation).

Full Local Loop Unbundling: The third party has total control over the local loop, in order to provide users both with voice and broadband services.

Local Loop Unbundling (LLU): The regulatory process of allowing multiple telecommunications operators the use of connections from the telephone exchange's central office to the customer's premises (the local loop).

Ladder of Investment: A dynamic regulatory policy aimed at favouring access of existing network infrastructures in the first place, and then inducing entrants to invest in their own access networks (climb the ladder) through an appropriate micro-management of the access terms.

Shared Access: The provision of access to the local loop by authorizing the use of the non-voice band frequency spectrum of the twisted metallic pair, whilst the local loop continues to be used by the incumbent to provide the telephone service to the public.

Static Efficiency: It occurs when marginal production costs are minimized (production efficiency) or when the price consumers pay in exchange of a good or service equals the production cost (allocative efficiency).

ENDNOTES

1. The paper has been circulated as working paper since 2001.
2. The equilateral triangle is the projection of the three-dimensional unit simplex onto the two dimensional Euclidean space. Each point's coordinates are found by computing the gravity center of mass points using the data entries as weights. Thus, the coordinates of a point P(a,b,c), a+b+c=1, are: P(b+(c/2),c * sqrt(3)/2.
3. In other words, all plots start at 1, then, for year 2006, the plot shows $\frac{P_{06}^{ULL} / P_{05}^{ULL}}{P_{06}^{ADSL} / P_{05}^{ADSL}}$ and, for year 2007, the plot shows $\frac{P_{07}^{ULL} / P_{05}^{ULL}}{P_{07}^{ADSL} / P_{05}^{ADSL}}$, where p^{ULL} is the monthly rental fee of an unbundled local loop and p^{ADSL} is the cheapest monthly fee of a broadband connection of the incumbent.
4. The reports are freely available at http://ec.europa.eu/information_society/policy/ecomm/library/ communications_reports /annualreports/previousyears/index_en.htm.
5. See http://www.oecd.org/sti/ict/broadband.

Chapter II
Is Regulation a Roadblock on the Information Highway?

James E. Prieger
Pepperdine University, USA

Daniel Heil
Pepperdine University, USA

ABSTRACT

Regulatory policy in telecommunications must balance short-term efficiency (low prices) against the firms' incentives to innovate, which have longer reaching impacts on economic welfare. Historically, policy tended to sacrifice dynamic efficiency for the sake of competitive prices and static efficiency. In the last few decades, economists and other researchers have begun to document the large welfare costs of ignoring dynamic efficiency. We analyze the theoretical impact regulation has on innovation. We then turn to the empirical evidence that regulation dampens firms' incentive to innovate in the telecommunications industry in general and the market for broadband Internet access in particular. Both product and process (cost reducing) innovation are discussed. The chapter forms a compendium of available research on the intersection of telecommunications regulation and innovation. We conclude with lessons the literature provides to policy makers and a discussion of future regulatory trends.

INTRODUCTION

For much of the 20th century, regulatory policy directed toward the telecommunications market was concerned with "getting the prices right". Regulators took the set of existing firms and products as given and sought prices that maximized consumer surplus, subject to the constraint that the regulated firm cover its costs of providing the current set of services. Although other regulatory objectives such as universal service played a role, regulation was framed within an essentially static view of the market. In the latter part of the century, however, as the pace of technological change increased in the telecommunications industry, it became clear that regulation could hinder innovation. In this chapter, we consider the evidence that regulation dampens firms' incentive to innovate. We begin by laying out the theoretical reasons underpinning this notion. In the main section of the chapter, we review the empirical studies in the literature, focusing on the U.S. market. In so doing, we find remarkably consistent evidence from numerous institutional and geographic settings that lighter forms of regulation encourage innovation.

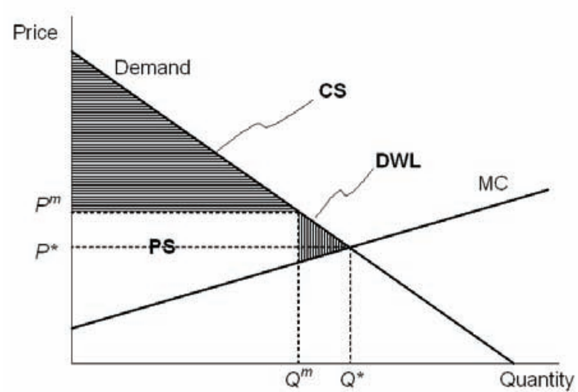
Before proceeding, we must ask what innovation is. The term is used in the economics literature to refer to everything from basic invention to new product introduction to diffusion of existing technology. We use the term in a broad sense to refer to the making available of something new in a given market. It is useful, however, to distinguish between process and product innovation. Process innovations are advancements in the methods of creating existing products, and may not be directly apparent to consumers. Process innovation lowers the cost of producing goods or services currently available to consumers. Product innovation is the creation (or diffusion to new markets, in our expansive definition) of new goods previously unavailable to consumers.

Regulation affects firms in many ways. Regulators historically deemed regulation justifiable, to reap the economies of scale and scope created by a single service provider and to further social goals such as universal service, while avoiding the inefficiency due to monopoly pricing. Regulatory control over prices and profits was chosen instead of reliance on competition or antitrust policy, which is generally not designed to prevent market expansion by firms with legitimate cost advantages over rivals. Competition law, at least in the U.S., furthermore does not outlaw the unilateral exercise of market power in setting prices.

The economic inefficiency created by prices above their competitive level is illustrated in Figure 1, which depicts the demand and marginal cost curves for a telecommunications service. Total economic benefits from the service are the entire area between the demand and the marginal cost curves. Benefits are maximized when quantity Q^* is provided (as would happen if the competitive price P^* is charged). These social benefits, shared between the consumers and the firm, arise because one part of society (consumers) consumes units for which it is willing to pay more than it costs another part of society (the firm) to produce. If a carrier with some market power charges a higher price, such as P^m, then a lower quantity Q^m of the service is purchased and the market loses benefits in the amount of area DWL in Figure 1. This deadweight loss, also known as the "Harberger triangle", represents the dollar value of the economic benefits lost to consumers and the firm from units between Q^m and Q^* that are not consumed at the higher price. Deadweight loss is a static loss in welfare, because it is calculated given the service is already available.

The Harberger triangle is not the end of the story, because a regulatory regime that attempts to squeeze static inefficiency out of market prices may create dynamic inefficiency. Dynamic efficiency stems from the additional net surplus created by new products and services. In Figure 1, the surplus obtained by consumers each period from the existence of the service is triangle CS, sometimes called the "Dupuit triangle".[1] The firm gains producer surplus, the unshaded area PS. If regulated prices are too low to encourage in-

Figure 1. Static versus dynamic efficiency



novation, so that the product is never introduced, then consumers and producers miss both benefits (although the firm saves the cost of innovating). Thus, as Bourreau and Doğan (2001) and many other authors point out, regulatory policy is a balancing act that often trades increased static efficiency for decreased dynamic efficiency. The focus of this chapter, the dynamic costs of regulation, has received much less attention than have the static effects (Joskow & Rose, 1989).

In the next section, we analyze the impact regulation has on innovation by means of a simple model (the details of which are in the appendix). We then turn to the empirical evidence on the interplay between regulation and innovation in the telecommunications industry in general, and the market for broadband Internet access in particular. In these sections, we include every relevant formal econometric study pertaining to the U.S. case that we could find, as well as representative studies using data from other countries or international comparisons.[2] Thus, the meat of this chapter—and our main objective—is a compendium of the available research on how innovation depends on regulation in the telecommunications industry.[3] We conclude with lessons the literature provides to policy makers and a discussion of future regulatory trends.

HOW DOES REGULATION AFFECT INNOVATION?

The likely impact of a market intervention can be analyzed by studying how the regulation changes firms' incentives. While some regulatory schemes such as price caps are known as "incentive regulation" in particular, it is important to realize that *any* regulation, if it affects the actions of the market participants at all, does so by changing their incentives. The economic study of regulation thus entails looking at how regulation changes the profit function of the firm. When the profit function changes, the actions the firm takes to maximize profit also change. Take the case of product innovation. Regulation can affect firm's incentives to product innovate through three channels. There are direct effects through mandates, indirect effects stemming from changes in the cost of bringing a new product to market, and indirect effects stemming from changes in operating profits gained by new products.

Most visible are the direct effects of regulation on innovation. For example, a line of business restriction placed on the U.S. Bell Operating Companies (BOCs) after the breakup of the AT&T Bell System in 1984 prevented the BOCs from manufacturing telecommunications equipment. The incentive to introduce a new manufactured product, therefore, was dwarfed by the cost to the firm of the legal difficulties that would have ensued. Other examples of direct impacts of regulation include social contracts between the regulator and the firm, which are often the outcome of a public utility commission's review of a merger case or renewal of a regulatory regime. In such contracts with the regulator, the firms often commit to putting new services on the market or making existing services available in more service areas.

The merger of Ameritech and SBC in 1999 provides an example of the latter. Ameritech, the incumbent local exchange carrier (ILEC) in five Middle Western states, had (in the eyes of the federal and state regulators) dragged its feet in making digital subscriber line (DSL) services available to customers. Before approving the merger, the Federal Communications Commission (FCC) extracted commitments from the companies to promote advanced services such as broadband Internet access. In particular, SBC was required to locate at least 10% of their advanced service facilities in low-income areas. The requirement had teeth: penalty payments of more than $2 billion were specified. Subsequent to the merger, SBC deployed DSL more rapidly in the area (Hu & Prieger, 2008).

In contrast to the obvious effect that direct regulatory prohibitions and deployment mandates have on product introduction, many (indeed, most) effects of regulatory policy are indirect. To see how, we introduce a simple model of the firm's decision problem, the details of which are in the appendix.[4] Any innovative firm has many potential ideas for new goods that it can develop and offer. Some projects would provide higher operating profit if implemented, others lower. The firm will introduce new goods only if they have operating profits in excess of innovation cost (which includes everything from basic research through development and product launch costs). The number of new goods passing the profitability threshold each period determines the rate of innovation. We can now use this framework to consider some indirect impacts of regulation on innovation.

The first type of indirect regulatory effect is on the cost side. Regulations that affect the cost of innovation include mandated regulatory filings and hearings before a public utility commission before introducing a new service. Cost studies that a firm must perform before a state commission approves rates for a new service are an example of the indirect costs of regulation. Hearings and cases in the legal and regulatory arena can also be expensive for the firm. For example, before (and even after) the BOCs could introduce information services there was a drawn-out string of federal court cases stretching from the late 1980's into the mid 1990's (Prieger, 2002).[5]

Such regulatory costs increase the cost of innovation. In response, some projects that were barely profitable before are lost—never brought to market—because of the increase in regulatory costs. However, highly profitable projects (which are typically so because they provide sizable benefits to consumers) are still pursued. Unless the additional costs created by regulation are large, or there any many potential projects that are minimally profitable before the cost increase, the number of new products lost because of these cost effects of regulation may be minimal.

The second type of indirect regulatory effect is on the benefit side. Regulations that effectively increase the time to market of a new product reduce the present value of the project by delaying the accrual of service revenue. Examples from telephone regulation abound. Many states and the FCC traditionally required hearings or minimum approval delays of tariffs for new services, designed to protect consumers from the "deleterious consequences of innovation," as one regulatory official put it (Oppenheim, 1991). The studies by Prieger (2001, 2002) cover examples from state and federal jurisdiction.

Regulation can also limit revenue from a new product through other means. One avenue is through competition. Under the unbundling regime instituted by the Telecommunications Act of 1996 in the U.S., incumbent telecommunications carriers must unbundle and lease to rivals certain parts of their network. The return on deploying the infrastructure is risky, and the carrier cannot fully recover the cost of investment after it is sunk. Thus, unbundling creates an asymmetry between incumbents (which bear the risk) and competitors (which, because they can stop renting at any time, do not). Renting unbundled network elements (UNEs) thus grants competitors a "free option", and the extra competition reduces the return on the incumbent's investment.[6] Gayle and Weisman (2007) note that while unbundling may readily increase competition in telecommunications, the increase comes at the expense of investment that is vital for future innovation.

Regulation can affect the competitive environment and the benefits a firm expects from introducing new products in other ways, as well. Regulatory policy can encourage competition through means other than mandating resale and unbundling, such as by requiring interconnection between competing networks. Regulation can also create legal monopoly (e.g., the Bell System in the U.S. until 1984), which enhances the incumbent's ability to appropriate the social benefit created by innovation. Monopolies, however, may cannibalize demand for one of their existing products by introducing a new one, which increases the opportunity cost of innovation (Arrow, 1962).

Other features of the general regulatory regime a carrier operates under can reduce the benefits from innovation. For example, "prudency reviews" were a common feature of rate of return regulation (RORR), the most common form of utility regulation in the U.S. for much of the 20th century (Kolbe & Tye, 1990). If the regulator deemed a failed investment to be imprudent, it would be stricken from the rate base upon which the firm's allowed rate of return was calculated, reducing the firm's revenue. RORR also attenuates incentives for process innovation, since excess returns gained thereby last only until the next regulatory review (Cabral & Riordan, 1989).[7] Alternative regulation, which includes price caps, earnings sharing plans, rate freezes, and other schemes may provide greater incentives for innovation. We discuss alternative forms of regulation more in the next section. For example, tariffs for new services require no cost studies or prior approval under the FCC's price cap regulation. Also, under price caps a dollar saved on cost (through process innovation) is retained by the firm, in contrast to RORR. However, even under alternative regulation the present value of the anticipated revenue from a new product is typically lessened, unless the regulation has no impact on the firm's activity at all.

When regulation reduces the revenues expected to follow from new products, fewer projects will pass the profitability threshold. Thus, similar to the case in which regulation increases the cost of innovation, when regulation limits revenue fewer new goods are introduced and the innovation rate drops.

In contrast to the usual view that regulation sacrifices dynamic efficiency on the altar of static efficiency, proponents of regulation sometimes claim that regulation encourages innovation. In the context of our simple framework, this could happen only if regulation lowered costs or increased revenue. It is difficult but not impossible to imagine cases where regulation accomplishes such changes. One oft-cited role for the regulator is to set technological standards in markets such as mobile telephony where coordination among firms, perhaps due to network effects in demand, is important. A standard imposed by the regulator that speeds consumer acceptance of a product can therefore increase revenue (or lower the cost of coordination). However, for every example of successful regulatory standard setting (e.g., perhaps, the GSM standard for European mobile telephony), one can find examples of failure. For example, the FCC's delayed approval of a standard and spectrum allocation rules for mobile telephony for about 14 years during the nascence of the industry in the U.S. Furthermore, industry-led efforts to coordinate are often successful absent regulatory intervention (e.g., the International Telecommunications Union (ITU) standards for Group 3 fax transmission).

Regulatory rules affecting competition, similarly, can also have conflicting effects on the rate of innovation. Increasing incentives for entrants to bring new services to market may diminish incentives for incumbents to do the same, and the net effect on innovation is ultimately an empirical matter in each market studied. We take up the empirical evidence in the next section.

EMPIRICAL EVIDENCE

Economic theory thus suggests that regulation may hamper innovation, at least by incumbents, but theory alone cannot accurately measure the impact in any given market. The numerous regulatory reforms in the U.S. telecommunications industry over the past few decades give researchers a unique opportunity to quantify the consequences of various types of regulation. Compared to the vast number of papers related to the static effects of various regulatory schemes, there has been only limited effort directed toward quantifying the impact of regulation on innovation. Joskow and Rose (1989) find this "distressing" given that the "static gains and losses from regulation are probably small compared to the historical gains in welfare resulting from innovation and productivity growth."

Nevertheless, some researchers have compared the amount of innovation under the traditional RORR and incentive regulation. In the following section, we outline the empirical findings concerning the amount of innovation under divergent regulatory schemes. We also examine attempts to quantify the effects of regulatory delay on innovation. In addition, we review empirical findings on how unbundled network elements (UNEs) affect innovation. Finally, we explore endeavors to measure the value of the loss to society from the postponement of product introductions due to regulation. Process and product innovation are covered in separate subsections, and a final subsection reviews the literature on the broadband Internet access market as an in-depth case study.

Before delving into the empirical literature, it is important to point out several possible pitfalls for researcher. Kridel, Sappington, and Weisman (1996) discuss the problems posed by demonstration effects, sequencing effects, and before-and-after study designs. The demonstration effect pitfall occurs when regulated firms, in an attempt to encourage favorable regulatory reforms, artificially "demonstrate" the success of these regulations. Thus, any positive action a firm takes after regulatory reform may not result from better incentives to innovate. On the contrary, the actions of the firm may merely be a strategic (and temporary) decision to encourage permanent regulatory changes that are favorable to the firm. If the demonstration effect is substantial, then the positive effects of incentive regulation that many researchers find may persist only in the short term.

A related hazard is the sequencing pitfall, in which firms hold off investment or innovation until the anticipated introduction of more favorable regulations. If the sequencing pitfall occurs, then innovations attributed to newly adopted regulations may also reflect only a short-term change in the firm's behavior. A third pitfall occurs with before-and-after empirical models (Sappington & Weisman, 1996), in which performance in a period before the new regulation is contrasted with a period after. Since regulatory change is not conducted under conditions of a controlled experiment, such models can confuse the effects of regulation with trends in innovation that are exogenous to the regulatory reform. With these potential pitfalls in mind, we can review the conclusions found in the empirical literature.

Process Innovation

Process innovation occurs when a firm, operating efficiently given its current technology, lowers its operating costs further by implementing new technology. Hence, researchers, when examining process innovation, often focus on changes in a firm's costs.[8] Stimulating process innovation is often a goal of regulatory reform. For example, one of the intentions behind the transition from RORR to incentive regulation was to provide firms with better incentives to reduce costs and deploy digital infrastructure.

The empirical literature concerning regulation's impact on firms' performance categorizes regulatory regimes differently, although generally a distinction is made between traditional RORR and alternatives. Ai and Sappington (2002) provide a good example of regulatory classification. They distinguish between RORR and three types of incentive regulations: rate case moratoria, earnings sharing regulations, and price cap regulations. Earnings sharing regulations give the firms greater control over which rates and services to offer, but require firms to return to consumers a percentage of their earnings above certain thresholds. Under most earnings sharing regulations, the firm is prohibited from increasing its earnings above a certain point. Firms with mandated earnings sharing will thus not have an incentive to lower costs beyond a certain level. Earnings sharing regulation is thus similar to RORR, in that both regulatory schemes focus on the profit being made by the firm rather than the prices of the services in question. Rate case moratoria are typically an intermediate regulatory scheme in which regulated firms are freed from traditional RORR and given greater control over setting rates. Under price cap regulation, there are no direct constraints on profit and the firm instead faces a limit on how high it can raise its prices. Greenstein, McMaster, and Spiller (1995) note that firms under most price-cap regulations have greater control over prices than have firms regulated by traditional profit regulation like RORR.

Kridel et al. (1996) review a variety of earlier papers related to incentive regulation and find mixed evidence concerning reductions in operating costs.[9] Ai and Sappington (2002) perform a careful study and test empirically whether the introduction of incentive regulation did in fact lower the cost of producing existing services. They observe that rate case moratoria do indeed lower production costs compared to RORR in their sample of ILECs in the period 1986-1999. Surprisingly, other forms of incentive regulation do not produce significantly lower production costs than RORR, holding all else constant. The authors suggest that this finding may be due to the regulated firms' fear that lower production costs will cause regulators to "ratchet" the standards for the firm's performance upwards, thereby making any gains temporary. When local competition is present, regulation does indeed reduce production costs relative to RORR. This finding supports the notion that incentive regulation should be complemented by increased competition to realize the theorized gains of incentive regulation.

The deployment of digital infrastructure under various regulatory schemes is an important aspect of process innovation since it lowers the cost of providing services. Moreover, increases in digital infrastructure enable new services that require greater digital capacity. Ai and Sappington (2002), Greenstein et al. (1995), Tardiff and Taylor (1993), and Taylor, Zarkadas, and Zona (1992) find that incentive regulation leads to increased deployment of digital infrastructure, such as fiber optic cables and digital switches, when compared to RORR. Greenstein et al. (1995) conclude that if all states had adopted some form of pricing regulation ILECs would have installed at least 75 percent more fiber optic cables than under the status quo. They do not, however, find evidence that earnings sharing would have produced results drastically dissimilar to those under RORR. Ai and Sappington (2002), in contrast, find that deployment of digital infrastructure is more extensive under all types of incentive regulation—including earnings sharing—than under RORR. Unlike the other studies, Ai and Sappington (2002) correct for the endogeneity of the regulatory regimes[10] in their econometric modeling, and thus have the strongest claim to finding true causal effects of regulation. They also find that the amount of digital infrastructure added to a local area under earnings sharing regulation increases when local competition increases.

To sum up the literature on process innovation: incentive regulation appears to spur the deployment of next-generation infrastructure, and perhaps to lower operating costs. The latter result does not always hold empirically, in contrast to the predictions from theory. The divergence may be explained by the fact that regulators in practice are often not able to commit to adhering to the incentive regulation when the temptation to return excess profit to consumers arises. If firms look forward to only short term gains from reducing costs, then their incentive to process innovate is blunted.

Product Innovation

Despite the limited amount of research on regulation and process innovation, until recently there has been more work looking at process innovation than at product innovation. Given the difficulty of counting innovations not created due to regulation, the imbalance is not surprising. Recent years have witnessed more attempts to quantify differences in product innovation under varying regulatory schemes, but the literature is still sparse. In addition to the usual comparison between RORR and incentive regulation, questions of regulatory delay become especially important when examining product innovation. Regulatory delay occurs when regulators prevent new products from entering the market until significant regulatory review has occurred. Regulatory delay may enhance social welfare by ensuring that new products meet certain guidelines, but may also create a disincentive for firms to release new services.

In an early attempt to study product innovation in the telecommunications industry, Mueller (1993) examined the effects of extreme deregulation. In 1987, while other states were adopting earnings sharing or price-cap regulation, lawmakers in Nebraska opted to remove nearly all restrictions on the telecommunications industry. Firms were allowed to introduce new products and change rates with little regulatory oversight. Of 100 new services offered by U.S. West that Mueller (1993) randomly selects for the study, 37 were first introduced in Nebraska. The result of Nebraska's experiment supports the contention that deregulation spurs firms to create new services. However, U.S. West likely opted to release new services in Nebraska first in order to demonstrate the benefits of such extreme deregulation. This is an example of the demonstration effects pitfall (Kridel et al., 1996), and it likely explains a portion of the increase in new services in Nebraska. Mueller's findings may have limited validity in other settings, especially since no formal econometric model is used in measuring the impact of extreme deregulation.

More recent attempts to quantify the amount of product innovation under various regulatory schemes have used more formal econometric models. Regulators in Indiana replaced traditional RORR with a mixture of incentive regulation and deregulation. Firms were allowed to set prices and the long regulatory delays witnessed under RORR were significantly reduced. Prieger (2001) estimates that the dominant ILEC (Ameritech) created new services 2 to 4.5 times faster than it did under the previous regime. Moreover, Ameritech, would have introduced up to twelve times as many services had reform been enacted at the beginning of the observed period. The author cautions, however, that the model consists of only two periods, one before introduction of the new regulations and one after introduction, with no "control" state. Thus the pitfall cited by Sappington and Weisman (1995), in which trends in overall innovation rates are indistinguishable from the actual regulatory effects, cannot be avoided.

Prieger (2004) confronts the problem posed by before-and-after study designs by using unique data covering three periods of regulation for information services offered by dominant ILECs. The first and third period had extensive FCC regulation that created significant regulatory delay and forced firms to file extensive paper work before the approval of new services. The second, interim period had lighter regulation. The empirical evidence shows that firms introduced considerably more new services during the interim than during the first or third periods. In fact, the model predicts that the rate of product innovation was anywhere from 60% to 99% higher during the interim. Furthermore, these new services reached consumers much quicker during the interim since firms did not face any significant regulatory delays.

The author's data and study design minimizes the potential for the before-and-after problem, but the sequencing pitfall remains a possible problem. If firms withheld innovations near the end of the first period in order to release them during the more profitable interim, then the significant increase in new services during the interim is not the result of better incentives to innovate under less regulation. The author addresses the sequencing pitfall and ultimately concludes that it does not significantly affect the results.

Prieger (2007) presents a theoretical model that predicts that a reduction in average regulatory delay would result in introduction of new services more quickly. These theoretical predictions are tested empirically by examining the number of new services offered in four states that adopted reforms designed to significantly reduce regulatory delay. The evidence confirms that the theoretical prediction: shorter regulatory delays lead to quicker product innovation. Thus, Mueller (1993) and Prieger (2001,2004,2007) all consistently find that lighter regulation does indeed encourage greater product innovation.

Increases in social welfare due to the creation of new products is an important aspect of the study of regulation and product innovation. While it is unrealistic to expect a precise measurement of the gains and losses to society caused by telecommunications regulation,[11] some attempts have been made to determine the welfare losses or gains from certain regulations.

Hausman (1997) examines regulatory delay by the FCC in approving the widespread availability of cellular telephones. He finds that the loss to consumer welfare in 1983 from regulatory delay was somewhere between $16.7 and $24.3 billion in 1983 dollars. Total losses were much higher, given that mobile telephony could have been introduced a decade earlier than it was in the U.S. In similar fashion, Hausman (1997) estimates that regulation preventing AT&T (before divestiture) and the BOCs (after) from offering voice mail services cost consumers $1.2 billion per year.

Prieger (2004), who examines the regulatory regime (CEI) put into place once the BOCs were allowed to offer information services, uses Hausman's (1997) calculations to estimate the effect of regulatory delay on voice mail services. The potential cost to consumers in delayed availability of voice mail services due to the CEI regime ranges from $690 to $910 million. Prieger (2004) notes that voice mail is merely one of dozens of information services delayed by regulatory action. However, Hausman's figures for voice mail cannot be extrapolated to other services, because many of the others were much less subscribed than voice mail services. Moreover, some of the services held up by regulatory delay were substitutes for a service already available. The introduction of such services would probably increase consumer surplus far less than would truly novel services. We have not found other attempts to quantify the loss in consumer welfare caused by regulation.

A Case Study of the Broadband Internet Access Market

We turn now to one specific sector of the telecommunications industry: the market for broadband Internet access. The spread of Internet access—first narrowband, and now broadband—is one of the most studied phenomena in the literature on regulation of telecommunications. As in the rest of the chapter, we focus mainly on the U.S., but also draw international studies into the discussion at times. Growth in broadband Internet access (hereafter, "broadband" for short) has been rapid in the U.S., especially compared to the spread of other recent services such as mobile telephony (Faulhaber, 2002). The growing importance of broadband in the national economy is large but difficult to quantify. The benefits of broadband that accrue to consumers are worth hundreds of billions of dollars per year in the U.S. (Crandall & Jackson, 2003). Total benefits are even higher, since business' profits are not included in the estimate. Such rapid growth raises questions related to policy. Did good regulatory policy in the U.S. encourage the spread of broadband? Or, as some claim (e.g., Hausman, 2002), could broadband have diffused even faster in the U.S. (as it did in other countries such as Japan and Korea) if regulatory roadblocks had been removed? In this section, we look at the evidence available.

As in other segments of the telecommunications industry in the U.S., regulatory policy toward broadband is a welter of partly coordinated (at best) state and federal efforts.[12] State level direct subsidies for broadband are relatively rare: one survey found that only three states targeted tax incentives toward broadband deployment in 2001 (Lee, 2001). However, not all states were included in the survey, and a later study found 15 states with broadband tax incentives (Wallsten, 2005). Wallsten (2005) estimates the impact of other state policies directed at broadband, including private-sector grants and loans targeted to deployment in underserved or rural areas and use of universal service mechanisms to stimulate investment. None of these is positively correlated with the per capita broadband rate in the state,[13] except for rural-targeted grants, and some even have negative correlation.

Federal subsidies for carriers are not available for broadband specifically, although rural and high-cost areas receive general support for infrastructure, some of which may enable advanced services. Federal demand-side subsidies include the "e-rate" for schools and libraries. Although billions have been spent on the subsidies, Flamm (2005) finds no measurable effect on broadband availability. In sum, while it is not hard to find case studies of this or that neighborhood, organization, or school that benefited from being brought online by a subsidy program, there is scant evidence that the state and federal money spent has had large enough impacts to be measurable by econometric studies. Some of the programs are relatively recent, however, and the conclusion may change as time elapses and more data become available.

Lee (2001) reports that at least 14 public utility commissions hold rate hearings for broadband rates, but Wallsten (2005) states that nowhere do states set rates for broadband, a discrepancy not easily reconciled. There is no federal rate regulation. In any event, we know of no study examining whether direct rate regulation of broadband affected its deployment by providers or take-up among consumers.

The likeliest places to look for the impact of state policy on broadband are the general regulatory scheme for telecommunications and the prices it allows incumbents to charge competitors for access to the local network. Both impacts occur through indirect channels. As described previously, RORR and price cap regulation lead to differing incentives to deploy new products—in this case, DSL. While the cable companies' decisions to offer broadband is not directly affected by public utility regulation in the U.S., any regulation affecting the deployment of DSL will indirectly affect the market for cable modem service, because they are substitutable to some degree.

Prieger and Lee (2008) examine broadband deployment data for the entire U.S. and find that areas under RORR have a lower probability of broadband availability than areas under price caps or rate moratoria. The impact of the form of regulation is not large, however. After controlling for which firm is the local incumbent, the presence of competition, and a host of demographic and economic characteristics of the area, price caps and rate freezes are each associated with an increase in the probability of broadband deployment in the postal code area of about one percentage point. Compare the increase with an average deployment of broadband in the ZIP code of 75% in 2000, the vintage of the data used. Broadband services are often not subject to either price caps or rate freezes even when basic telecommunications services are. So, it is perhaps not surprising that the evidence is consistent with rate of return regulation (in which revenue from all sources is typically regulated) dampening the incentive to deploy new services compared to alternative regulation. The correlation found in Prieger and Lee (2008) may not be causal for a host of the usual reasons, as they discuss. Regulatory regimes are not randomly assigned, alternative regulation may be offered to companies in exchange for commitments to roll out advanced services, and companies favored with alternative regulation may wish to "demonstrate" its beneficial effect to the regulator by speeding broadband deployment.

The other way states may indirectly affect broadband is through policy toward UNE rates for the parts of the local network. States do not set UNE rates unless negotiations between the private parties break down, but the threat of state rate setting affects the relative negotiating positions of the players. Pindyck (2007) notes that competitive local exchange carriers (CLECs) have a significant advantage in bargaining. If the parties are unable to agree on a rate, then the state regulator will typically impose a relatively low price to ensure CLECs can effectively compete in the market. The effect of UNE rates on broadband is indirect, because an ILEC's DSL packet-switching facilities are not subject to unbundling. However, competitors wishing to offer DSL to subscribers without duplicating the local network could also purchase the "last mile" segment between the incumbent's wire center and the subscriber's premises as a UNE. Since DSL does not require the whole line, until 2003 competitors could "line share" with the incumbent by renting just the high-frequency part of the local loop as a UNE to offer DSL. These various forms of unbundling enable competition in DSL. Cable companies, in contrast, have never been required to open their networks to competitors.

The empirical literature on the impact of unbundling on performance and competition is sizable, and here we cover only those empirical studies examining broadband innovation.[14] The impact of unbundling policies directed toward a subset of the industry may not be as important in general as competition itself between cable and telephone companies. Howell (2002) and Maldoom, Marsden, Sidak, and Singer (2003) use sets of national case studies to find that unbundling is less successful than intermodal competition at speeding broadband deployment in developed nations.[15] In another international comparison, Wallsten (2006) finds no relationship between full-loop unbundling requirements and broadband penetration, but that sub-loop unbundling is correlated with lower penetration.[16] García-Murillo and Gabel (2003) likewise find no impact of unbundling requirements on broadband availability within a country or on the percentage of population with broadband access. Relying solely on data from the U.S., Wallsten (2005) concludes that where more UNEs are rented, lower broadband penetration results. Curiously, he finds the opposite correlation for lines that are resold instead of purchased as unbundled elements, which casts doubt upon a purely causal explanation for either result.[17] The sole study we found that associates unbundling with increased broadband access is García-Murillo (2005), which uses a small international cross-section of countries.[18]

Should we expect high or low UNE rates to stimulate broadband? Hausman (2002) argues forcefully that allowing competitors to rent facilities after they are deployed by the incumbents causes the incumbents to invest less in infrastructure. He attributes the early lead of cable modem service over DSL to the former's closed networks. In Hausman's (2002) view, low UNE rates retard the spread of broadband, at least among incumbent carriers. In accord with this notion, the rate of DSL subscription growth rose markedly after the FCC's line-sharing rule was lifted in 2003 (Hazlett & Bazelon, 2005), although it is impossible to assess from a simple before-and-after comparison whether the change is entirely causal. Burnstein and Aron (2003) use state broadband subscription rates to indicate that lower UNE rates discourage broadband, although their estimate is not statistically significant.

On the other hand, lower UNE rates encourage entry by competitors, thus spurring competition that may spill over to the broadband market as well (García-Murillo & Gabel, 2003; Ford & Spiwak, 2004). Partially supporting this conclusion, Distaso, Lupi, and Manenti (2006), using data from Europe, and Prieger and Lee (2008), using U.S. data, find that areas with lower UNE rates are correlated with more broadband availability, but the sizes of the effects are small (although statistically significant). The latter study has the largest sample size and number of control variables of any, and furthermore controls for all state-level variables that do not vary over time in the econometric model. Further investigation reveals that the effects of UNE rates on broadband are greatest in states with alternative regulation. Thus, although the evidence in the literature is mixed, perhaps the tentative conclusion at this date is that UNE rates have little measurable impact on broadband deployment.

CONCLUSION AND OUTLOOK

For policymakers attempting to improve current regulatory schemes, the limited research available on innovation presents a problem. Each regulatory setting is unique, and thus presents unique incentives to participating firms. The external validity of the case studies is unknown, and not all instances of regulation have been thoroughly studied. What then can we conclude? Fortunately, a consensus exists in the economic literature and some general lessons are clear. In short, heavier regulation does place roadblocks on the information highway.

Incentive regulation appears to spur process and product innovation. Whether examining total deregulation (Mueller, 1993), incentive regulation (Prieger, 2001), or regulatory delay (Hausman, 1997; Prieger 2001,2002,2007), studies typically find a negative relationship between the number of roadblocks created by government regulation and the amount of product innovation. While some studies may be picking up a demonstration effect in part, as long as the regulator periodically reviews the firms' performance under the lighter regulatory regime and maintains the threat of re-instituting heavier-handed regulation (as was done in Indiana a few years past, for example), the improved innovation should continue. However, the regulator must avoid the temptation to use periodic reviews to confiscate excess returns created by process innovation, for then dynamic efficiency is lost in the name of short-term, static welfare gains.

Innovation and dynamic efficiency, however, are not the only goals of regulators. On the contrary, regulators are often instructed by legislation to strive toward competing goals, such as efficiency and universal service, and must find a balance. Rather than reading our review as necessarily calling for complete deregulation, policymakers should instead treat it as pointing out some of the dynamic costs of regulation of which they may not have been adequately aware.

As regulators attempt to balance competing factors, however, they will find that the large potential costs from lost innovation often tip the scales in the direction of lighter regulation. These costs are higher in times of rapid technological change, such as the industry now finds itself in. The trend in telecommunications is toward convergence of voice, data, and video communication. Regulatory schemes that discriminate between the types of information communicated (e.g., FCC distinctions between information and telephone services) or the mode of carriage (e.g., cable systems vs. the telephone network vs. the Internet) based on the arbitrary historical accretion of rules are sure to fail to promote dynamic efficiency to its greatest extent. For example, when voice communication via VoIP (voice over Internet protocol) is carried over the cable system's network on one end to the Internet and terminated on the telephone network on the other end, is it a telephone call? Or is it just transmission of data? A more appropriate question: why should it matter? Rules that force regulators to split increasingly fine hairs to categorize services and providers are bound to hamper investment in new technology and services and to favor less efficient technology or carriers over more efficient in some cases.

Banerjee et al. (2007) discuss principles for future regulation to be guided toward the goal of dynamic efficiency. In the era of rapid technological change and convergence, good regulation consists of reversing the regulator's past emphasis on static efficiency to the detriment of dynamic efficiency. Instead of trying to force prices down to their long-run competitive levels immediately, or attempting to pick (and thus artificially creating) technological winners through asymmetric regulation, dynamically efficient regulation seeks to remove entry barriers wherever possible to allow unhindered intermodal competition. As technology such as the Internet erodes the monopoly bottlenecks remaining in the telecommunications network, intermodal competition becomes ever more feasible. However, removal of barriers need not be carried forward into creating "negative entry barriers" by providing artificial advantages to intramodal competitors (as some claim the unbundling rules have done for wireline telephony in the U.S. [Hausman, 2002]). Promoting inefficient entry at the expense of incumbents does not enhance dynamic efficiency and innovation. The challenge for regulators in the future will be to "do no harm" as they seek to level the playing field to let all participants compete, without imposing *a priori* notions of neutrality of outcomes that stifle innovation. Without entry barriers in the market, dynamic efficiency will help solve the problem of static efficiency as competition moves prices toward their efficient level.

REFERENCES

Ai, C., & Sappington, D. E. M. (2002). The impact of state incentive regulation on the U.S. telecommunications industry. *Journal of Regulatory Economics, 22*(2), 133-159.

Arrow, K. (1962). Economic welfare and the allocation of resources for inventions. In R. Nelson (Ed.), *The Rate and Direction of Inventive Activity* (pp. 609-625). Princeton, NJ: Princeton University Press.

Banerjee, A., Madden, G., & Tan, J. (2007, January) *Regulatory trends: New enabling environment*. Paper presented at the ITU New Initiatives Programme workshop on "The Future of Voice", Geneva. Retrieved March 30, 2008, from http://www.itu.int/osg/spu/ni/voice/papers/FoV-Madden-Banerjee-Tan-Draft.pdf

Bourreau, M., & Doğan, P. (2001). Regulation and innovation in the telecommunications industry. *Telecommunications Policy, 25*(3), 167-184.

Burnstein, D. E., & Aron, D. J. (2003). Broadband adoption in the United States: An empirical analysis. In A. L. Shampine (Ed.), *Down to the wire: Studies in the diffusion and regulation of telecommunications technologies (pp. 119-128).* New York: Nova Science Publishers.

Cabral, L. M. B., & Riordan, M. H. (1989). Incentives for cost reduction under price cap regulation. *Journal of Regulatory Economics, 1*(2), 93-102.

Crandall, R. W., & Jackson, C. L. (2003). The $500 billion opportunity: The potential economic benefit of widespread diffusion of broadband internet access. In A. L. Shampine (Ed.), *Down to the wire: Studies in the diffusion and regulation of telecommunications technologies* (pp. 155-193). New York: Nova Science Publishers.

Distaso, W., Lupi, P., & Manenti, F. (2006). Platform competition and broadband uptake: Theory and empirical evidence from the European Union. *Information Economics and Policy, 18*(1), 87-106.

Dixit, A.K., & Pindyck, R.S. (1994). *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Faulhaber, G. R. (2002). Broadband deployment: Is policy in the way? In R. W. Crandall & J. H. Alleman (Eds.), *Broadband: Should we regulate high-speed Internet access?* (pp.223-244). Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

Flamm, K. (2005, February). *The role of economics, demographics, and state policy in broadband* *availability.* Paper presented at the PURC/London Business School Conference on "The Future of Broadband: Wired and Wireless, 2005", Gainesville, FL.

Ford, G. S., & Spiwak, L. J. (2004). *The positive effects of unbundling on broadband deployment* (Policy paper No. 19). Washington, DC: Phoenix Center for Advanced Legal & Economic Public Policy Studies. Retrieved March 30, 2008, from http://www.phoenix-center.org/pcpp/PCPP19Final.pdf

García-Murillo, M. (2005). International broadband deployment: The impact of unbundling. *Communications & Strategies, 57,* 83-105.

García-Murillo, M., & Gabel, D. (2003, September). *International broadband deployment: The impact of unbundling.* Paper presented at the 31st Telecommunications Policy Research Conference, Washington, DC.

Gayle, P., & Weisman, D. (2007). Efficiency tradeoffs in the design of competition policy for the telecommunications industry. *Review of Network Economics, 6*(3), 321–341.

Greenstein, S., McMaster, S., & Spiller, P. (1995). The effect of incentive regulation of infrastructure modernization: Local exchange companies' deployment of digital technology. *Journal of Economics and Management Strategy, 4*(2), 187–236.

Hausman, J. (1997). Valuing the effect of regulation on new services in telecommunications. *Brookings Papers on Economic Activity, Microeconomics*, 1-38.

Hausman, J. (2002). Internet-related services: The results of asymmetric regulation. In R. W. Crandall & J. H. Alleman (Eds.), *Broadband: Should we regulate high-speed Internet access?* (pp.129-156). Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

Hausman, J., & Sidak, G. (2005). Did mandatory unbundling achieve its purpose? Empirical evidence from five countries. *Journal of Competition Law and Economics, 1*(1), 173–245.

Hazlett, T., & Bazelon, C. (2005). *Regulated unbundling of telecommunications networks: A stepping stone to facilities-based competition?* Paper presented at the 33rd Telecommunications Policy Research Conference, Arlington, VA.

Howell, B. (2002). *Broadband uptake and infrastructure regulation: Evidence from the OECD countries* (Working Paper BH01/01). Wellington: New Zealand Institute for the Study of Competition and Regulation.

Hu, W.-M., & Prieger, J. (2008). The timing of broadband provision: The role of competition and demographics. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission* (pp. 241-259). Hershey, PA: IGI Global.

Joskow, P., & Rose, N. (1989). The effects of economic regulation. In R. Schmalensee and R. Willig (Eds.), *Handbook of Industrial Organization, 2,* 1449-1506. Amsterdam: North-Holland.

Kahn, A., Tardiff, T., & Weisman, D. L. (1999). The Telecommunications Act at three years: An economic evaluation of its implementation by the Federal Communications Commission. *Information Economics and Policy, 11*(4), 319-365.

Kolbe, A. L., & Tye, W. B. (1990). The Duquesne opinion: How much "Hope" is there for investors in regulated firms? *Yale Journal on Regulation, 8*(1), 113-157.

Kridel, D., Sappington, D. E. M., & Weisman, D. L. (1996). The effects of incentive regulation in the telecommunications industry: A survey. *Journal of Regulatory Economics, 9*(3), 269-306.

Lee, C. H. (2001). *State regulatory commission treatment of advanced services: Results of a survey.* Columbus, OH: National Regulatory Research Institute.

Lyon, T. P., & Huang, H. (1995). Asymmetric regulation and incentives for innovation. *Industrial and Corporate Change, 4*(4), 769-776.

Maldoom, D., Marsden, R., Sidak, J. G., & Singer, H. J. (2003). *Competition in broadband provision and its implications for regulatory policy.* Retrieved March 30, 2008, from http://ssrn.com/abstract=463041

Mueller, M. L. (1993). *Telephone companies in paradise: A case study in telecommunications deregulation.* New Brunswick, NJ: Transaction Publishers.

Oppenheim, J. (1991). In B. G. Cole (Ed.), *After the breakup: Assessing the new post-AT&T divestiture era,* (pp. 310-315). New York: Columbia University Press.

Pindyck, R. (2007). Mandatory unbundling and irreversible investment in telecom networks. *Review of Network Economics, 6*(3), 274-298.

Prieger, J. E. (2001). Telecommunications regulation and new services: A case study at the state level. *Journal of Regulatory Economics, 20*(3), 285-305.

Prieger, J. E. (2002). A model for regulated product innovation and introduction with application to telecommunications. *Applied Economics Letters, 9*(10), 625-629.

Prieger, J. E. (2004). Regulation, innovation, and the introduction of new telecommunications services. *Review of Economics and Statistics, 84*(4), 704-715.

Prieger, J. E. (2007). Regulatory delay and the timing of product innovation. *The International Journal of Industrial Organization, 25*(2), 219-236.

Prieger, J. E. (2008). Product innovation, signaling, and endogenous regulatory delay. *Journal of Regulatory Economics, 34*(2), 95-118.

Prieger, J. E., & Lee, S. (2008). Regulation and the deployment of broadband. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission* (pp. 241-259). Hershey, PA: IGI Global.

Resende, M. (2000). Regulatory regimes and efficiency in US local telephony. *Oxford Economic Papers*, *52*, 447-470.

Riordan, M. H. (1992). Regulation and preemptive technology adoption. *RAND Journal of Economics*, *23*(3), 334-349.

Sappington, D. E. M., & Weisman D. L. (1996). Potential pitfalls in empirical investigations of the effects of incentive regulation plans in the telecommunications industry. *Information Economics and Policy, 8*(2), 125-140.

Tardiff, T., & Taylor, W. E. (1993). *Telephone company performance under alternative forms of regulation in the U.S.* National Economic Research Associates.

Taylor, W. E., Zarkadas, C. J., & Zona J. D. (1992, June). *Incentive regulation and the diffusion of new technology in telecommunications.* Paper presented at Ninth Annual Conference, International Telecommunications Society, Sophia-Antipolis, France.

Wallsten, S. (2005). *Broadband penetration: An empirical analysis of state and federal policies* (Working Paper 05-12). Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

Wallsten, S. (2006). *Broadband and unbundling regulations in OECD countries* (Working Paper 06-16). Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

WEFA Group (1995). *Economic impact of deregulating U.S. communications industries.* Bala Cynwyd, PA: The WEFA Group.

Willig, R. (1976). Consumer's surplus without apology. *American Economic Review, 66*(4), 589-597.

KEY TERMS

BOC: Bell Operating Company. The BOCs in the U.S. were the large ILECs created when AT&T was split up in 1984.

CLEC: Competitive Local Exchange Carrier. A CLEC is a public telephone company that provides local telecommunications service in competition with the ILEC.

Consumer Surplus: The benefits for buyers created by consumption of a good or service, net of the purchase price, denominated in currency units. Consumer surplus is created when a unit is consumed at a price less than the maximum willingness to pay for it.

DSL: Digital Subscriber Line, a technology that provides a dedicated digital circuit between the local telephone company's central office and an end-user's premises, allowing broadband data transmission over existing telephone lines.

Deadweight Loss: potential economic surplus that is not realized by any participant in a market.

ILEC: Incumbent Local Exchange Carrier. A public telephone company in the U.S. that provides local telecommunications service.

Process Innovation: Innovation that lowers costs through the creation or adoption of new technology.

Product Innovation: Innovation that makes new products or services available to consumers.

RORR: Rate of Return Regulation. The most common historical form of telephone regulation in the U.S., in which the return on capital invested by the firm is limited to a "fair" rate.

UNE: Unbundled Network Element. UNEs allow CLECs to offer competitively priced alternatives to the incumbent's services rapidly. UNEs are purchased at wholesale rates from the incumbent telephone companies. UNEs include the local loop, switch port, switching, transport, signaling systems and databases (such as operations support systems and directory assistance). UNE-Platform (UNE-P, generally no longer available in the U.S.) combined all elements necessary to offer local service.

ENDNOTES

1 Areas under the demand curve, such as the Harberger and Dupuit triangles, are only approximations of consumer welfare when there are income effects in demand. However, as long as the income effects are not too strong, they are good approximations in many cases (Willig, 1976).

2 Given the paucity of formally peer-reviewed, published research, we cast a wider net than is often seen in such literature reviews. The reader is cautioned that some of the working papers and other unpublished studies we cite may not have undergone rigorous peer review. This is particularly true of the section on the broadband market below.

3 We have undoubtedly missed some research, and welcome readers bringing omissions to the attention of the first author.

4 For more sophisticated modeling of a regulated firm's incentives to innovate, see Riordan (1992), Lyon and Huang (1995), and Prieger (2007,2008).

5 The cases revolved around the FCC's *Computer III* series of orders, stretching from the *California I* decision rendered in June 1990 until the *California III* decision of October 1994, when *Computer III* was remanded.

6 The term "option" comes from the real options literature, in which the option value of delaying risky investment is priced into the firm decision problem (Dixit and Pindyck,

1994). The adjective "free" does not mean that the competitors bear no costs of entry using UNEs, but instead that they are not required to pay for the risk reduction that UNEs offer them. See Hausman (2002) for the argument applied to DSL infrastructure.

7 Kahn, Tardiff, and Weisman (1999) argue, to the contrary, that RORR encourages innovation by reducing the risks involved, since the firm is guaranteed a specified return on R&D.

8 Incentive regulation may reduce costs by means other than process innovation, however. The famous Averch-Johnson effect, for example, maintains that a firm choosing its input mix under RORR will not minimize costs given the output produced and its current technology.

9 See also work by Resende (2000), who estimates a yearly "efficiency score" (a measure of how productively a firm uses inputs) for U.S. ILECs. He then regresses the scores on regulatory variables to find that alternative regulation is robustly and significantly correlated with better efficiency.

10 Endogeneity of an explanatory variable occurs when it is correlated with the econometric error term, which violates a fundamental assumption of ordinary least squares regression. For example, if incentive regulation is more likely to be adopted in poorer performing areas, then incentive regulation may be correlated with worse outcomes in a regression, even though the true causal impact is in the other direction. Ai and Sappington (2002) discuss the issue thoroughly and provide a solution.

11 Nevertheless, see WEFA Group (1995) for a bold attempt.

12 This section draws on Prieger and Lee (2008).

13 Subscription rates reflect both innovation (in the sense of diffusion: where the service is available) and other factors, such as the usual supply and demand considerations for existing markets.

14 Much of the literature on UNEs looks at whether unbundling leads to investment in infrastructure by incumbents and facilities-based entry by competitors. See also Hausman and Sidak (2005), who discuss arguments for whether unbundling leads to more innovation. They conclude unbundling did not lead to innovation in the five cases they study.

15 See also Distaso et al. (2006) for further evidence that intermodal competition is an important driver of broadband penetration.

16 In full loop unbundling, the entire "last mile" of copper between the wire center and the subscriber is rented to the competitor. Sub-loop unbundling entails renting access to only the last part of the "last mile," which in the case of DSL can allow superior transmission performance.

17 Federal regulations require that any service that the incumbent local exchange company offers to retail customers has to be offered to CLECs at wholesale prices. These rules are distinct from the unbundling regime and UNE rates.

18 In the estimation where García-Murillo (2005) finds a statistically significant positive effect of unbundling on broadband usage, there are 12 variables in the regression but only 18 observations, resulting in a near-perfect fit ($R^2 = 0.98$).

19 We treats costs and revenue as known quantities here. If they are uncertain, there is no change needed in our discussion if the firm is risk neutral. If the firm is risk averse, regulation has additional effects on the firm's decision problem (see, e.g., Prieger, 2007).

20 In mathematical terms, the effect on the amount of innovation of a(n infinitesimally) small increase in c is $M'(c) = -m(c)$.

APPENDIX

Here we explicate the model, alluded to in the text, of how regulation indirectly affects innovation. Consider the panoply of new goods that a regulated firm could potentially offer. Potential goods differ in the amount of operating profit they will provide.[19] Let the density *m* of operating profits *R* describe the distribution of potential outcomes, so that $M(x) = \int_x^\infty m(R)dR$ is the number of projects that earn more than *x*. Assume in our simple model that the innovation cost *c*, which includes everything from basic research through development and product launch costs, is the same for all projects. Then the amount of innovation in which the firm chooses to engage is *M*(*c*), for only projects to the right of *c* are profitable (see Figure 2).

The extra costs that regulation creates for innovators (examples of which are given in the chapter) increase *c*. The effect on the amount of the firm's innovative activity is clear: as *c* shifts to the right in Figure 2, there are fewer worthwhile projects to pursue, *M*(*c*) decreases, and there is less product innovation. However, it is important to note that if regulation induces only a small change in *c*, then there will be only a small effect on innovation. Only those products that were barely profitable that are lost because of regulatory costs. How many such projects are not pursued depends on the density of projects near *c*.[20]

Regulation also can affect the benefit side of an innovator's calculations. In this chapter we give many examples of how regulation reduces the revenues expected to follow from new products. In terms of the model, the density curve for operating profit shifts left (perhaps also changing shape) when this happens

The net effect of regulation's influence on the costs and benefits of introducing new products is shown in Figure 3. First, consider costs. As regulation increases costs from c_0 to c_1, the mass of innovation that becomes unprofitable is depicted by the horizontally striped area. Next, note that the vertically striped area is the mass of new products that become unprofitable due to the leftward shift of operating profit curve *M*. It can be seen that changes in the revenue of potential services may have a larger effect on the amount of innovative activity of the firm than changes in cost. Changes in cost affect only the marginally profitable services, whereas changes in revenue affect the mass of all formerly profitable services.

Figure 2. The determination of the amount of innovation

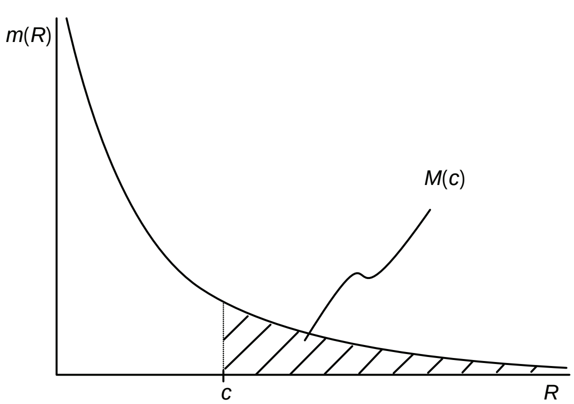
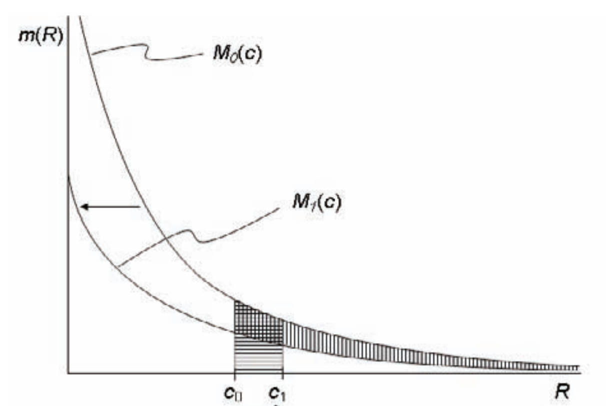


Figure 3. Changes in the amount of innovation due to regulation



Chapter III
Reforms in Spectrum Management Policy

Claudio Feijóo
Universidad Politécnica de Madrid, Spain

José Luis Gómez-Barroso
Universidad Nacional de Educación a Distancia (UNED), Spain

Asunción Mochón
Universidad Nacional de Educación a Distancia (UNED), Spain

ABSTRACT

Wireless/mobile technologies are absolutely necessary for any agent who wants to participate in the new information and communication technologies market playfield. The means for the deployment of the wireless/mobile technologies is the radio spectrum. Thus, a rational assignment and an efficient use of radio spectrum become sine qua non condition for the sector development. The objective of this chapter is to present the reforms in the radioelectric spectrum management mechanisms that are currently being drafted (or that are even being applied) as well as to assess their advisability and timeliness. In particular, the chapter assesses the three deepest changes that are being considered: authorisation of the secondary market, usage of auctions for primary allocation and full liberalisation of spectrum usage.

INTRODUCTION

As users' demand for access to contents and applications moves towards a scenario characterised by mobility and ubiquity, the aim of integrating the provision of services and business models into a single structure becomes increasingly both necessary and attractive. This is the idea that lies behind the concept of *Next Generation Networks* (NGN). Very simply speaking, a NGN is a single network that delivers multiple applications (voice, data, video) to multiple devices, whether fixed or mobile.

Therefore, the deployment of the NGNs is the technical and business element around which the future evolution of the information and communication sector revolves. This upcoming scenario will be characterised by the fading of the boundaries between previously separate markets, because no access technology by itself (at least with the technical conditions expected as of today) could present the optimal characteristics for satisfying all the requirements involved in the NGN concept. In particular, wireless/mobile technologies are absolutely necessary for any agent who want to create (or to update) technological platforms capable of thriving in the new playfield.

The means for the deployment of the wireless/mobile technologies is the radio spectrum. Thus, a rational assignment and an efficient use of radio spectrum become *sine qua non* condition for the sector development. The management of spectrum (the set of rules that govern the conditions to access it, the requisites for its usage and the rights it entitles) will define the characteristics and sustainability of future information and communications technologies (ICT) markets.

Moreover, the consequences of the decisions about spectrum planning and management go beyond the sector. The innovation introduced in ICT markets regarding new services and applications hugely depends on the strategy followed by the operators which is reliant on market conditions. As just said, those strategies are conditioned by the spectrum regulatory framework. The knock-on effects, undoubtedly, will affect general users and, particularly, business organisations for which communications services are becoming a key strategic asset, no matter the sector of economic activity.

The objective of this chapter is to present the reforms in the radioelectric spectrum management mechanisms that are currently being drafted (or that are even being applied) as well as to assess their advisability and timeliness.

The paper is structured into six sections. Section 2 details what could be considered as the "traditional spectrum management model". Next, the reasons that make advisable a change of model are analysed and the most outstanding features of the transformation are summarised. Sections 4 and 6 describe and assess the three deepest changes that are being considered: authorisation of the secondary market, usage of auctions for primary allocation and full liberalisation of spectrum usage. The paper ends with the conclusions that result from the previous analysis as well as with a few personal recommendations serving as a guide for change.

THE TRADITIONAL SPECTRUM MANAGEMENT MODEL

The traditional regulation of the radioelectric spectrum is conditioned by two assumptions: the radioelectric spectrum is a scarce resource (or at least a limited one) and it is also a valued good which is used in services that are essential to society: applications connected to security, defence and response to emergencies, telecommunications and radio broadcast, transportation, scientific research, etc.

With these assumptions, the radioelectric spectrum is considered as a public domain good by practically all administrations worldwide. As a consequence, the procedures for granting the right to using it are strictly regulated. Basically, after planning the usage given to each frequency band and specifying the service that can use it and the necessary technology, the right of usage is granted to certain agents through administrative decisions detailing the obligations of the licensee .

This system resolves the problems that may rise due to interference. Additionally, since the first step of the process, planning, is coordinated in supranational organisations, an acceptable (not always complete) global compatibility is guaranteed of the communication systems which also generates important economies of scale in the electronic device market.

As compensation for the rights of usage granted to the licensees, they are usually expected to pay an annual fee (fixed or subject to their results) and, frequently, to meet other commitments (usually regarding network deployment or level of coverage).

THE NEED FOR A CHANGE OF MODEL

Objectives for Spectrum Management

The main objective of any spectrum management mechanism is (or should be) reaching the greatest usage efficiency to the extent possible. This general principle of efficiency can be broken down into three dimensions: technical efficiency, economic efficiency, and social efficiency. Note that these concepts are applied to the allocating and management mechanisms, and not to the usage made by the agents of the spectrum they have been granted.

- Technical efficiency requires that the greater amount of frequencies be available to be used effectively in the provision of services[1]. This implies that, should there exist several technologies with which to offer a specific service under the same conditions, the one that uses the less resources (frequencies) should be the preferred option. Additionally, the guard bands should be reduced as technically feasible: that is, without this restriction leading to the interference between the different radioelectric signals making it impossible to make good use of certain bands.
- An economically efficient spectrum allocation is one in which there is no way to reallocate spectrum such that everyone is better off, all else being equal (Morris, 2005). That is, the economic efficiency would be maximum when, from the combination of possible uses of the resource (the spectrum), the one valid for "producing" the services optimising the usefulness of the agents present in the market was chosen. Thus, allocating spectrum to services and/or technologies that are not the most demanded by the market should be avoided. Also, the rights of usage should be removed for those that have become obsolete.
- Social efficiency implies that the radioelectric spectrum should be used in those services or applications maximising the wellbeing of society. In this sense, certain bands of the spectrum must be allocated to activities such as research or security, activities that would be discarded if only the maximisation of business benefits (or even the usefulness of the consumer) were to be considered in the short term.

In all cases, the analysis should be extended by introducing a prediction of future development or, in other words, optimising a dynamic definition and not a simply static one, of the efficiency. Also worth noting is the interrelation that exists between the three dimensions. For example, an increase in the technical efficiency releases frequencies, thus allowing the entry of new agents or the introduction of new services; the resulting increase in competition results in benefits for the users or the society as a whole.

Characteristics of the New Scenario

The gradual introduction of improvements and innovations both in the technologies for handling the radioelectric spectrum and in those that use it have caused deep changes in the markets. In turn, the changes in the markets have promoted research which, in a sort of virtuous circle, has boosted new technological changes.

Although no causality can be established between the latter and former advances, what can indeed be considered is which of these transformations affect each of the three dimensions of efficiency in the usage of the spectrum:

- As regards the technical efficiency:
    - Development of technologies that reduce the amount of spectrum required for sending the same amount

of information. A paradigmatic and contemporary case is the transition from analog to digital format in terrestrial television broadcast.
  - o Introduction of dynamic spectrum access techniques (such as software defined radio or cognitive radio) allowing highly flexible management and (re)use2 of spectrum across all its dimensions (frequency, time, location and code) (refer to Nekovee, 2006).
  - o Appearance of technologies that use a very broad range of spectrum with, simultaneously, a very low emission of power (ultrawideband technologies), which causes a very small level of interference and its possible "coexistence" with other services operating on the same band.
  - o Innovations with repercussion in the access part of the networks (mesh networks, femtocells) allowing the reuse of frequencies and an increase of the efficiency in their usage.
- As regards the economic efficiency:
  - o Proliferation of services that require radioelectric spectrum for operating.
  - o Existence of services that have become obsolete and that use much more spectrum than would be necessary with other technological alternatives or, in the most extreme case, that do not use at all the frequency bands allocated (and that, as a consequence, should be reallocated).
  - o Growing usage of bands that do not require usage license, thus playing down the need for such a detailed planning. Not only growing, but successful as well: two examples are WiFi and Bluetooth.
  - o Fast evolution of the technology (not always predictable), thus hindering the *a priori* allocation of spectrum bands to specific technologies which can become rapidly obsolete.
- As regards the social efficiency:
  - o Strategic importance of many services that require spectrum and that are considered to be tools that promote the development of the information society.

Keys for a New Model

As opposed to these new circumstances, the traditional model appears as excessively complex and lacking flexibility. This fact has led to the appearance of a current of opinion favourable to the introduction of market economy criteria in radioelectric spectrum management[3]. In fact, many national authorities in charge of spectrum management[4], the European commission[5] or the ITU[6], have released documents supporting the modification of the current spectrum management models and defending the inclusion of market mechanisms. Moreover, the reform of the regulating framework is already effective in several countries.

The new ways of managing the spectrum intend to increase flexibility and transparency, as well as the speed of the response to technological innovations. The actual demand and the value given by the market to each band of frequencies are new governing criteria. Obviously, any change should improve global resource usage efficiency, while maintaining certain inalienable technical requirements.

There are three possible paths leading to change. These are listed in ascending order in the process of allocating the spectrum usage rights:

- Modifications in the conditions of the allocated license: relaxation of some clauses but, especially, authorisation of the transmission of rights (secondary spectrum trade).
- Modifications in the license-allocation mechanisms: usage of auctions.

- Modifications in the definition of the licenses: avoiding to bind the license to specific technologies (technological neutrality) or even to specific services (service neutrality).

These three paths do not represent different alternatives. Several of these tools can be chosen, and this could be done simultaneously as well as in succession. However, they are not completely independent, and as we will see below, the introduction of certain changes is almost unavoidably connected with other modifications of the model.

Below we will analyse the characteristics of these eventual reforms.

SECONDARY SPECTRUM MARKET

Concept

"Trading the spectrum" represents the total or partial subrogation of the rights that have been awarded to an agent to make use of a segment of the public radioelectric domain.

Trading creates a new way of accessing the spectrum. Interested companies can obtain the rights of usage through the primary allocation, if they are chosen by the appropriate authority responsible for spectrum allocation, or through the secondary market, purchasing the rights of usage from another agent.

The reasons that would lead an operator holding a usage license over the public radioelectric domain to subrogate it would be of an economic nature: said operator receives a remuneration from the agent acquiring the right (directly or indirectly, if the substitution is connected to other agreements between the two parties) or the transmission represents a cost reduction (for example, avoiding to have to pay the annual radioelectric fee or tax).

On the other hand, the interest of the operator receiving the rights of usage for the spectrum can be quite varied: accessing a new market, acquiring additional spectrum in order to improve the provision of their services, avoiding the entry of new competitors, etc.

Modes of Transferring the Usage Rights

Table 1 establishes a classification of the different criteria with which a right transfer could be carried out:

Figure 1. Spectrum management evolution (Source: Prepared by the author)

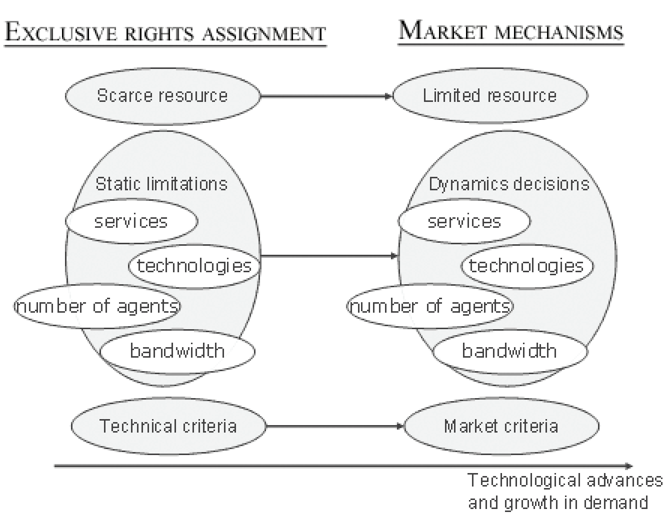


Table 1. Modalities of usage rights transfer (Source: Prepared by the author)

Criterion	Modalities	Peculiarities
Temporary	Permanent	The subrogation is established for an unlimited time, with the granting agent permanently relinquishing the right of usage (irrevocable transfer)
	Provisional	The subrogation is established for a period of time defined in the agreement
Geographical	Total	The granting agent relinquishes the right of usage of the spectrum in the whole geographical scope of the license
	Regional	The granting agent and the grantee share a part or all the frequency bands in different geographical areas
Frequency bands	Complete	The grantee of the rights of usage has to subrogate in the rights and obligations of the granting agent
	Partial	The granting agent maintains the obligations taken on before the Administration. It is essential to define the obligations acquired by the grantee, if any
Revocability	Revocable	According to the conditions stipulated, the granting agent can revoke the right transfer
	Irrevocable	The granting agent permanently relinquishes the right of usage of the spectrum

All the categories in Table 1 imply the relinquishment of the granting operator of all or part of their rights. In the so-called right "sharing", the spectrum is used by two operators without this representing a substantial reduction of the conditions of usage of the first holder. The conditions for a shared usage of the spectrum have their limit in the "supportable interference" or maximum level of noise generated by the new user that does not disturb the service of the original band holder. This modality is tightly related to the new possibilities offered by the technological progress mentioned previously, such as the ultrawideband or intelligent radio techniques.

Advantages and Disadvantages Resulting from Trading the Spectrum

Opening the radioelectric spectrum to the market can provide a series of advantages:

- Increase of the technical efficiency in the usage of the spectrum: the agents will have incentives to make a more efficient usage of the spectrum they have been granted, minimizing the amount of spectrum they require for providing their services and putting the rest on the market.
- Increase in competition: the increase of the technical efficiency represents indirectly the increase of the available spectrum and, as a consequence, of the amount of service offering agents.
- Promotion of innovation: the establishment of a free exchange system allows for technological innovations to reach the end user earlier, since on the one hand the companies with rights of usage of the spectrum have incentives to introduce said innovations and, on the other hand, the new operators have a mechanism to enter the markets which is much faster and flexible than the primary frequency allocation.

As opposed to the potential advantages sketched out previously, the secondary spectrum market poses some problems which should be considered before determining the spectrum granting procedure:

- Generation of uncertainties: the grant of the spectrum modifies the competition conditions established implicitly in the primary spectrum allocations, where the total number of licenses were specified, providing the market with a certain degree of stability which favoured the commitment of investments.

- Distortion of the competition: certain agents can make the most of the frequency granting to accumulate spectrum and thus achieve positions of market domination. Additionally, agents could appear that operate with merely speculative intentions, rising the price of the spectrum without providing any value neither to its management nor to its efficient usage.
- Appearance of other market failures: failures regarding the concepts of social and technical efficiency (conflicts due to interference, difficulties for the development of standards), and also due to the costs of transaction. For Valletti (2001), "one should not only question how serious these failures are, but also if the regulator would be able to find better solutions than the market".
- Inefficiency in the usage of the resource: an excessive fragmentation of the spectrum could originate a significant increase of the unavailable frequencies as they would be used as guard bands. Additionally, international coordination problems could appear, or merely regional ones, should the frequency grant be geographically limited.

MODIFICATION OF THE PRIMARY ALLOCATION METHOD: THE AUCTION

Characteristics of the Spectrum Auctions

Tenders have been the predominant mechanism (although not the only one[7]) for primary spectrum allocation. The tender allows the regulator to determine through economic criteria, among others (also social, even technical), which applicant is better qualified to use the spectrum.

Despite the aforementioned, the usage of auctions is not unknown in the sector. Licenses have been auctioned for years and in a great deal of countries. However, there was a point of inflection in the consideration of auctions as a spectrum allocation mechanism. During 2000 and 2001, several European countries auctioned third generation (3G or UMTS) mobile-communications licenses. The amount reached by the bids[8] even led the spectrum auctions to make the headlines in the media.

Since then, research on auctions started proliferating, attempting to reach the optimum allocation mechanisms for the spectrum in terms of efficiency as well as of revenue maximization because of the fact that the design of the auction model fully conditions the results obtained from the award process (Banks et al., 2003; Milgrom, 2004; Plot and Salmon 2004).

When there are multiple objects to be sold, like spectrum licenses, the seller has many options to consider. The first decision the agency in charge of managing the spectrum must take is whether to sell the objects separately in multiple auctions or jointly in a single auction.

- The main formats if licenses are sold in different auctions are[9]:
  - Sequential auctions: Licenses are sold one at a time in separate auctions.
  - Simultaneous auctions: Groups of related licenses are sold at the same time in different auctions and bidders can bid on any of the items. All auctions would close at the same time when no new bids are received in any license for a period of time. The Simultaneous Multiple-Round (SMR) Auctions (also called the Simultaneous Ascending Auction) have discrete, successive rounds and participants submit a se-

ries of single-item bids for the desired licenses. Then, the high bids for each license become the standing prices for the next round of bidding. The auction ends when no new bids are posted in a round in any license.

- If the seller chooses to sell all the licenses at a time in a single auction (especially when the items are not completely homogeneous), they must face the question about how to package the goods offered for sale. The usual standard consists in allowing the bidders to choose the packages for themselves. This type of auctions are known as combinatorial or package auctions, and among them, the following stand out (Cramton et al., 2006):
  - Iterative combinatorial auctions: At the beginning of each round, bidders submit prices for one or more packages. Then the auctioneer solves the winner determination problem, announces provisional allocation and asks prices for the next round. The auction ends at a fixed deadline or when no new bids are submitted.
  - Ascending proxy auctions: Bidders report their valuation about packages to a proxy agent. The proxy agent bids on the bidder's behalf in each round. The auctioneer considers all bids (from present and past rounds) and chooses his best feasible collection of bids (at the most, one bid from each bidder). The auctioneer announces provisional allocation and the process continues until no new bids are submitted.
  - Clock-proxy auction: Hybrid auction format that begins with a clock phase and ends with a final proxy round. In the clock phase, the auctioneer announces prices for each item being sold and bidders indicate the demanded quantities. Prices will continue going up until there are no items with excess demand. Then, the auction finishes with a proxy round in which bidders submit their preferences to a proxy agent. This auction format combines the advantages of the clock phase (simple and transparent price discovery) and the proxy phase (efficiency and revenue maximization), mitigating the exposure problem[10].

Of these possible auction formats, the combinatorial ones are particularly appropriate if there are strong complementarities between licenses (synergies[11]) as well as substitution possibilities. These auctions allow the participants to bid both for complete packages (all-or-nothing bids) as well as for individual licenses. However, and despite these advantages, their usage is not yet generalised due to their complexity both for the auctioneer (in the solution of the winner determination problem) and for the bidders when placing their bids.

In fact, the Federal Communications Commission (FCC), which has been conducting auctions of licenses for electromagnetic spectrum since 1994, uses SMR auctions[12]. As explained by Cramton et al. (2006), despite the SMR auction cannot be considered a combinatorial auction, most SMR auctions allow bidders to withdraw bids, mitigating the exposure problem and providing the bidders with the option of bidding for the licenses they are most interested in. Nevertheless, recently the FCC has also included the possibility of modifying the SMR auction allowing combinatorial or package bidding in some auctions.

Comparison Between Tender and Auction

Below are analysed a series of items that appear in any debate between supporters of one or another method of allocation. However, it should be stated that the characteristics of the market in question and, especially, the design of the process (be it auction or tender) are basic parameters that could alter the general conclusions. A well planned tender is always preferable to a deficient auctionm, and vice-versa. The objective would be thus to analyse which is the preferable method, assuming a correct design in both options.

- Tenders are accused of lacking transparency and credibility. Indeed, regardless of the rigour of the assessment criteria, the selection procedures include a subjective element, which is why they are sometimes known as beauty contests. This subjectivity usually rises the suspicion that the organisations in charge of making the decisions could carry out their obligations in a biased manner. Occasionally, the decisions are criticised, and even contested, thus affecting the credibility of the license awarding process and, generally, that of the responsible administration.

  Auctions, on the other hand, represent a transparent and objective way of awarding the licenses.
- As an argument against the auction, some adduce that the tender allows to allocate the spectrum to those operators with greater investment commitments, thus affecting the extension and quality of service or the creation of jobs.

  However, an appropriate prequalification procedure (including coverage and quality criteria) guarantees that the awardees of the auction will have the technical and financial capabilities required to quickly initiate service provision.

  Tenders lower the barriers of entry to the market for operators, since they do not request any prior payment in order to start operating (or, if they do so, the amount is much smaller). As a consequence, operators can carry out a rational distribution of the investment and have initially more economic resources to deploy infrastructures and develop services.

  The same thesis can be turned around in the case of the auction: the expenditure required to win an auction can also be considered as an incentive for the fast expansion of the networks. In fact, this is the only way in which the winner of an auction can recover the investment (sunken cost) that represents the price paid for the license.
- Related to the prior argument, some object that winning operators in an auction will assign the cost of the license directly to the end services they offer. Or, in other words, that consumers will pay higher prices. It has not been proven whether this behaviour is possible with operators. In his study, Bauer (2003) concludes that "despite weak hints that license fees influence prices, no statistically significant evidence for lasting effects could be detected".
- The tender creates economic income associated to the power of market created by limiting *ex ante* the number of agents present in the market (Hazlett, 2003).

  Nothing changes in the case of the auction, unless the license obtained is allowed to be transferred.
- A frequently used argument in favour of the spectrum auctions is that they provide the greatest income possible for the use of a public resource. This income could go to social plans in the industry itself (for example, financing the universal service).

  However, this is not an argument with absolute validity. Apart from the fact, of course, that collusive behaviours (where the bidders agree the price beforehand) must be avoided in the auction, the truth is that the income depends on the auction model and of the valuations and strategies of the bidders. Additionally, the result of the auction is always a mystery which would lead an administration with "aversion to risks" to prefer knowing beforehand what they would take and would chose the tender. In these cases, the type of service is also a parameter for decision: in services with low demand, there could be less bidders than auctioned rights.

- Exogenous factors may appear (an economic depression, a new technology) that make difficult honouring the commitments taken on at the time of awarding a tender.
  It is not infrequent for deployment or service requirements to be included in auctions. In that case, nothing would be different from a tender. It is true that the relaxation or suppression of the obligations connected to the awarding of a license is linked to the use of the auction since it leaves the tender without parameters of decision.

LIBERALISATION OF THE SPECTRUM USAGE

As stated in section 2, the allocation of a specific spectrum usage license involves obligations that are traditionally very strict: service to be provided, specific technology that must be used, limitations on the power of emission, geographic scope where broadcasts are allowed and bandwidth of each channel. Spectrum usage liberalisation allows for some or even all the aforementioned conditions to be relaxed or omitted, in a way that the agents that enjoy the rights over the public radioelectric domain are given the capability of deciding the usage and technology that they will implement.

Technological neutrality (free choice of the technology with which to provide the service) does not present too many problems, both from the conceptual point of view and in practice, particularly if the technological alternatives, as is usually the case, do not present radically different emission characteristics.

More complex (understood as more “revolutionary”) is the configuration of a spectrum regulating model also allowing service neutrality (free choice of usage given to the band). There are basically three models for achieving this “full” liberalisation of the spectrum:

- In the “private property” model, the contents and duration[13] of the usage rights are defined and a primary allocation of the latter is carried out. Once the licenses have been awarded, the agent holding the right has complete freedom of action[14], with the only restrictions of higher-level regulations (commercial legislation or competition defence legislation).
- In the *commons* model, it is considered as a community resource and, thus, there are no particular usage rights defined[15]. Not even a formal license is required to operate. This is a modality which has already been applied to certain “free usage” bands, such as those used by WiFi.
- As an alternative between the two previous models, Faulhaber and Farber (2003) propose a private property model avoidable under conditions of non-interference. In this regime, the spectrum could be shared and sold without any more restrictions in the usage than those referred to the power of the emission. However, this right of usage would be preferential and non-absolute since any other agent would be entitled to emit on that band provided they met the condition of non-interference. It would be a mechanism similar to the commons one described in section 4.2 but without the need for a prior agreement. Thus, a usage regime would be established *de facto* which could known as a “conditioned *commons*”.

The liberalisation of the spectrum usage faces, in any of its modalities, the problems that the traditional method of thorough allocation of services had resolved: global compatibility of communication systems, creation of standards (with the subsequent scale economies in the electronic device market) and prevention of interference.

The problem of interference is particularly acute in the case of the commons, and quite particularly if the usage is intensive and the resource becomes scarce ("tragedy of the commons"). The other two models resolve (at least partially) the problem but at the expense of increasing (maybe even extraordinarily) the transaction costs.

Although the private actors become the stars in these models, there remains a role for the spectrum management organisations: monitoring that the users meet the public radioelectric domain protection conditions, controlling the exposure limitations for health protection, risk assessment for radioelectric emissions, controlling that negotiations between agents are led under a competition framework. As a rule of thumb, public *ex ante* intervention is minimised and *ex post* control increases.

CONCLUSION

That the radioelectric spectrum needs a new management model (or at least a renewed one) is a consensus among public actors with interest in the market and researchers. Considering the analysis carried out in this chapter (and particularly the advantages and disadvantages presented in each case), a series of guidelines are proposed for the introduction of reforms in the spectrum management mechanisms:

- The reform must necessarily be progressive in depth (what measures), scope (what frequency band) and timetable. Regardless of its appropriateness, a full and immediate liberalisation of all the commercially useful spectrum is not feasible.

  This implies a progressive replacement of the *ex ante* regulation of the traditional scheme by a relaxed regulation (or a completely *ex post* one). The period of cohabitation between different models shall be long or even indefinite, since full liberalisation should not necessarily be the end objective.
- The modifications in the spectrum management policy shall have significant effects on the market only in the medium-term. It is critical to design a scenario which, starting from the current situation, identifies the objectives (basically in terms of efficiency). It shall be necessary to periodically monitor the degree of achievement of the objectives, underlining the best practices and introducing the necessary corrective measures.
- The assessment should precede any reform, assuming the enormous complexity of the current occupation in each band. The analysis shall condition to a great extent the specific measures that may be applied. Particularly, it will be useful to select a series of pilot bands on which to test alternative models.
- Given the different characteristics presented by the different bands it may not be feasible to apply a single spectrum management model. As a consequence, it would be necessary to carry out an analysis on a band-by-band basis determining the basic management model that is more adequate for each one (of among the possible options of "traditional regulation", "progressive introduction of market mechanisms" and "liberalised model"), and identifying the different services their usage can be dedicated to.
- The introduction of market mechanisms is compatible with fixing a series of technical and/or economic efficiency criteria useful for the control that the governmental and spectrum management organisations must exert.

  Thus, resource usage thresholds could be established in order to penalise those agents operating below the minimum level of efficiency, or, as a last resort and in cases of continuous inefficient usage, to revoke their rights of usage. This type of tool would also be useful in fighting against anticompetition practices such as the accumulation of spectrum in order to establish entry barriers to the market.

- All the processes regarding spectrum management must be public. This implies the publication of the global strategy of the organisation in charge of the spectrum, the updated maintenance of the frequency record (including relevant information on the allocations: frequency, bandwidth, duration, date on which the allocation was granted and expiry date, service, possibility of secondary business, price) and information on the degree of usage of each band, on the efficiency criteria and on the conflict resolution procedures. Specifically, the frequency record is important for an effective introduction of the spectrum trade (refer to Bykowsky, 2003).
- In less general terms, and considering the specific situation of certain services or bands, the following corrective measures could be considered:
    - The introduction of trading mechanisms should start from some "simple" bands, such as those dedicated to radiolinks. The experience gained here would later be applied to more complex bands for their offer–demand dynamics.
    - Given the rise of some technologies that operate in common usage/license-less bands (such as WiFi or Bluetooth), it would be necessary to investigate whether increasing the range of the spectrum dedicated to this type of usage would encourage the development of innovative applications.
    - The "digital dividend" resulting from the television "analog blackout" represents a great opportunity for introducing some criteria attempting to increase the efficiency in the usage of the spectrum.
    - The third generation (and similar or succesive) mobile communications present an opportunity to provide the broadband access to to those regions or zones that are still devoid of this type of connections. The spectrum currently allocated to this service seems to be insufficient and operators should be given the opportunity to access new bands (be it "traditionally", be it via the secondary trade).

REFERENCES

Ausubel, L. M., Cramton, P., McAfee, P., & McMillan, J. (1997). Synergies in wireless telephony: Evidence from the broadband PCS auctions. *Journal of Economics and Management Strategy*, 6(3), 497-527.

Banks, J., Olson, M., Porter, D., Rassenti, S., & Smith, V. (2003). Theory, experimental and the Federal Communications Commission spectrum auctions. *Journal of Economic Behaviour and Organization*., *51*(3), 303-350.

Bauer, J. M. (2003). Impact of license fees on the prices of mobile voice services. *Telecommunications Policy*, 27(5-6), 417-434.

Bykowsky, M. (2003). A secondary market for the trading of spectrum: promoting market liquidity. *Telecommunications Policy*, 27(7), 533-541.

Coase, R. (1959). The Federal Communication Commission. *The Journal of Law & Economics*, *2*, 1-40.

Cramton, P. (2002). Spectrum auctions. In Cave, M., Majumdar, S. & Vogelsang, I. (Eds.), *Handbook of Telecommunications Economics* (pp. 605-639). Amsterdam: Elsevier Science.

Cramton, P., Shoham, Y., & Steinberg, R. (2006). *Combinatorial auctions*. Cambridge: The MIT Press.

Falch, M., & Tadayoni, R. (2004). Economic versus technical approaches to frequency management. *Telecommunications Policy*, *28*(2), 197-211.

Faulhaber, G. R., & Farber, D. J. (2003). Spectrum management: property rights, markets, and the Commons. In Cranor, L. F. & Wildman, S. S. (Eds.), *Rethinking rights and regulations: Institutional responses to new communication technologies* (pp. 193-226). Cambridge: The MIT Press.

Hazlett, T. W. (2003). Liberalizing US spectrum allocation, *Telecommunications Policy*, *27*(7), 485-499.

Klemperer, P. (2002). How (not) to run auctions: The European 3G telecom auctions, *European Economic Review*, *46*(4-5), 829-845.

Marx, L. M. (2006). Economics at the Federal Communications Commission, *Review of Industrial Organization*, *29*(4), 349-368.

Milgrom, P. (2004). *Putting auction theory to work*. Cambridge: Cambridge University Press.

Morris, A. C. (2005). Spectrum auctions: Distortionary input tax or efficient revenue instrument? *Telecommunications Policy*, *29*(9-10), 687-709.

Nekovee, M. (2006). Dynamic spectrum access – Concepts and future architectures. *BT Technology Journal*, *24*(2), 111-116.

Plot, C. R., & Salmon, T. C. (2004). The simultaneous ascending auction: Dynamics of price adjustment in experiments and in the UK 3G spectrum auction. *Journal of Economic Behaviour and Organization*, *53*(3), 353-383.

Spiller, P. T. & Cardilli, C. (1999). Towards a property rights approach to communication spectrum. *Yale Journal on Regulation*, *16*(1), 53-84.

Valetti, T. M. (2001). Spectrum trading. *Telecommunications Policy*, *25*(10/11), 655-670.

Weiss, T. A. & Jondral, F. K. (2004). Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency. *IEEE*. *Communications Magazine*, *42*(3), S8-S14.

Werbach, K. (2004). Supercommons: Toward a unified theory of wireless communication. *Texas Law Review*, *82*(4), 863-973.

White, L. J. (2000). 'Propertyzing' the electromagnetic spectrum: why it's important, and how to begin. *Media Law & Policy*, *9*(1), 29-30.

KEY TERMS

Allocation: Allocation of a specific band of frequencies for their use by one or more services (terrestrial or spatial radiocommunication, or by the radio astronomy service) under specified conditions.

Primary Spectrum Allocation: Allocation of rights of usage of the spectrum by the appropriate spectrum management authority.

Frequency Band: Fraction of frequencies of the radioelectric spectrum that present similar characteristics as regards the propagation of the electromagnetic waves.

Free Usage Band: Frequency band, the use of which is not subject to the possession of a license awarding usage rights.

Secondary Spectrum Market: Market where the holders of spectrum usage rights transfer said rights to other agents.

Technological Neutrality: Characteristic of frequency allocation processes specifying the service for which a specific band is to be used, but not the specific technology to be used for providing it.

Service Neutrality: Characteristic of frequency allocation processes where the service (and thus, the technology) to which a specific band is to be allocated is not specified.

Radioelectric Spectrum: Limited natural resource made up by all the radioelectric waves (in the frequencies between 9 KHz and 3000 GHz) that propagates through space without the need for an artificial guide.

ENDNOTES

1 Which, in economic terminology, would represent being at the limits of the production possibilities (of resource usage).

2 Consider that extensive spectrum usage measurements show that considerable parts of the spectrum, although dedicated to specific services, are actually not used for significant periods of time, ranging from seconds to minutes (Weiss and Jondral, 2004).

3 Actually, it is a "reinforcement" and not an "appearance" of this current of opinion. As far as 1959, Coase already argued that the spectrum should be seen just as another production factor for which the value must be determined on the free market, while the interference problem should be solved by the regular private property right legislation (Coase, 1959).

4 As an example in Europe, that of the British regulator OFCOM. Refer to the document "Spectrum framework review: Implementation plan" released in 2005 (available at http://www.ofcom.org.uk/consult/condocs/sfrip/).

On the other hand, the FCC quotes as their first objective in their Web site: "advance spectrum reform by developing and implementing market-oriented allocation and assignment reform policies". Refer to http://www.fcc.gov/spectrum/

5 "Policy aims to facilitate spectrum access across the EU through market mechanisms". Communication of the Commission "i2010 – A European Information Society for growth and employment" COM (2005) 229 final. Brussels, 1.6.2005.

6 Refer to, as an example, the ITU document "Market mechanisms for spectrum management". A document that summarizes the results of a workshop with the same title held in Geneva, Switzerland, from 22 to 23 January 2007 (available at http://www.itu.int/osg/spu/stn/spectrum/workshop_proceedings/STN.MMSM-2007-PDF-E.pdf).

7 Drawing of lots represents a fast, economic and transparent method for choosing from several prequalified applicants. Despite draws were used by the FCC until the nineties, they do not represent an efficient solution, since the winner of the licenses may not be the one who values them the most or who offers the most socially beneficial service and deployment plan.

8 Measured by the income obtained by the Government, not all auctions had similar success. Actually, the results were extremely different from some countries to others, oscillating between the 20 euros per capita of Switzerland and the 650 euros per capita obtained in the UK auctions. This major dispersion was basically due to the poor auction design in some countries and to the time sequence in which the auctions were carried out (Klemperer, 2002).

9 Cramton (2002) states the advantages and disadvantages of these auction models with open and sealed bids, in spectrum allocation.

10 In the presence of synergies, bidders are affected by the exposure problem, as they can earn negative profits if they fail to win complementary objects.

11 Synergies among licenses are classified according to Ausubel et al. (1997) as local and global synergies. Local synergies are those gains in value that specifically arise

from obtaining two or more geographically neighbouring licenses. Global synergies are those gains in value which accrue from obtaining increased numbers of licenses or markets: economies of scale or scope among multiple licenses which arise irrespective of their geographic locations.

12 The Simultaneous Multiple-Round auctions fit in environments where bidders need to be able to shift their bidding among substitutes and to bid simultaneously on licenses that they would like to acquire together (Marx, 2006).

13 Although it would be possible for the Government to effectively transmit the property of the right over the spectrum, this does not seem to be the most probable option (refer to the defence of this option in Spiller and Cardilli, 1999, or in White, 2000). It would be expected that it was done for a specific number of years (even with the possibility of renewal) which, legally, would take the figure closer to a possession or to a right of usage and enjoyment.

14 This would include the right to do business with the license. As stated in section 3, the reforms in the management of the spectrum are not completely independent and it does not seem conceivable in a framework where the usage liberalisation is introduced without the secondary market.

15 “A universal communication privilege, allowing anyone to transmit anywhere, at any time, and in any way, should be the baseline rule for wireless communication” (Werbach, 2004).

Chapter IV
Next Generation Access Networks and their Regulatory Implications

Ricardo Gonçalves
Universidade Católica Portuguesa (Porto), Portugal

Álvaro Nascimento
Universidade Católica Portuguesa (Porto), Portugal

ABSTRACT

The deployment of Next Generation Access networks (NGAs) is likely to have a significant impact on the telecommunications' value chain and, consequently, on the necessary regulatory remedies. In particular, unlike existing telephone networks, NGAs present regulators with a dilemma, in so far as the possibility of regulatory intervention after network deployment (negatively) affects investment incentives for such a deployment. We review the current discussion surrounding NGAs and discuss some of the main regulatory challenges it presents.

INTRODUCTION

Next Generation Networks are slowly but increasingly being used at the core of the public switched telephone network (PSTN). A Next Generation Network (NGN) is, according to the ITU, "a packet-based network able to provide services including Telecommunication Services", i.e. a multi-service (voice, data, video or other media) network based on packet-switching (as opposed to the traditional circuit-switching of the PSTN networks). But whilst the core network has been (or will be) replaced by a Next Generation Network, able to carry all those services, the access network has lagged slightly behind.

As network operators deploy (or consider deploying) such Next Generation Access networks (NGAs), several interesting questions have to be raised. From an operator's perspective, it is important to understand how NGAs affect the value chain, for instance, by uncoupling service provision from specific network infrastructure requirements. In particular, vertical integration in the presence of NGAs appears to be a less sensible strategy than it is under existing telephone networks.

From a regulatory perspective, NGAs are likely to be an economic bottleneck and, as such, regulatory intervention may be warranted. However, a dilemma presents itself: unlike current telephone networks, NGAs do not exist yet. It is natural to expect that the possibility of regulatory intervention after network deployment (negatively) affects operators' incentives to deploy them in the first place.

This paper is a survey of NGAs and their main regulatory implications. OFCOM, the UK telecommunications regulator, is (arguably) the institution which has given this topic more thought. Hence, it should not be surprising that we refer to its documents extensively. Additionally, the paper was written with a view to increasing awareness of the importance of NGAs in the telecommunications sector and, inevitably, to the regulatory challenges it presents. As such, we deliberately refrain from advocating a specific course of action, as we believe the discussion is still at the beginning.

WHAT IS A NEXT GENERATION ACCESS NETWORK?

An access network, commonly referred to as the local loop, is the part of the telecommunications network which connects the subscriber to the local exchange, where switching equipment is installed. Traditionally, access networks consisted of copper wire connections used solely for telephone services. Recently, these copper wire connections have proved to be good carriers for additional services such as broadband connections.

Most broadband connections provided by telephone operators are xDSL connections. xDSL is a broadband technology which uses existing copper wired telephone networks to carry data traffic. The provision of xDSL services requires the network operator to install additional equipment in an exchange, so that voice traffic is separated from data traffic (they travel together along the copper-based access network). Data traffic is sent to the DSLAM (Digital Subscriber Line Access Multiplexer), the entry point in the IP network, which concentrates subscriber lines and their respective traffic.

By far the most popular type of xDSL is ADSL, which provides more capacity downstream (from the network to the subscriber) than it does upstream (from the subscriber to the network). The problem with xDSL is that not all existing copper wired subscribers can benefit from it in the same way. The analysis of several xDSL products, such as ADSL2, ADSL2+, VDSL, HDSL[1] or SDSL[2], highlights a trade-off between maximum download speed and local loop length. For some types of xDSL (e.g. VDSL), that length is very small (up to 1.5Km), but can deliver maximum download (52Mbps) and upload (16Mbps) speeds which are much faster than ADSL. These speeds make the provision of triple play services – three communications services (voice, data and television) over a single broadband connection – a real possibility. For instance, OFCOM (2007) suggests that 25Mbps would be sufficient to carry simultaneous multiple HDTV[3] channels, broadband internet and voice services.

Copper-based networks are investing in ADSL2+ deployment, which allows speeds up to 24Mbps. However, several factors such as the length, quality and dimensions of the copper cable and various other noise-contributing factors limit the overall availability of such speeds[4]. This suggests that offering triple play services through ADSL2+ may face significant limitations.

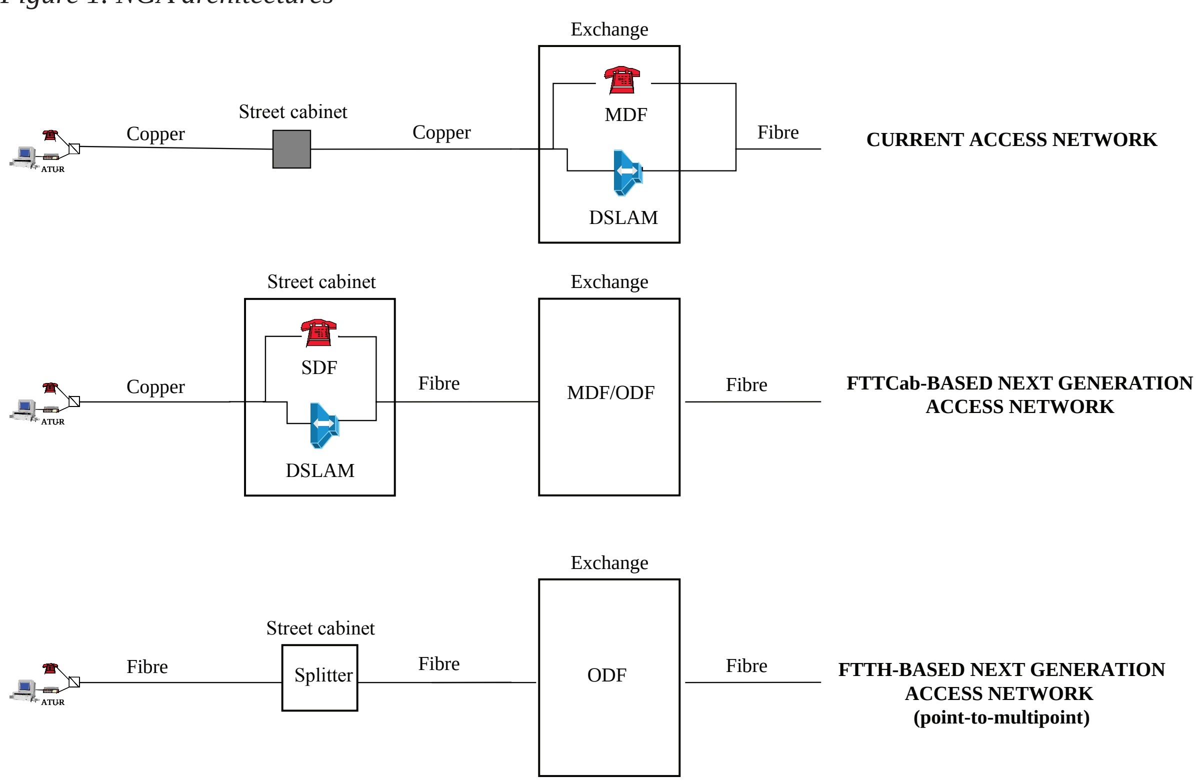
There does not appear to exist, as of yet, a good definition of a NGA – such a definition should, as expected, combine the various service characteristics (speed, symmetry, quality).

OFCOM (2006a) defines next generation access as "broadband access services that are capable of delivering sustained bandwidths significantly in excess of those currently widely available using existing local access infrastructures or technologies" (p. 10). This definition merges two concepts: (i) the maximum bandwidth available (headline speed) and (ii) the continuous availability of current peak speeds (approx. 20Mbps) so as to allow triple play.

A definition based solely on speed is probably not the most correct approach. It is important to note that in order to provide a sustainable bit rate, improvements in the upper layers of the network may be necessary. In fact, OFCOM (2007) raises the question as to whether the main short to medium term bottleneck may be the limitations of the backhaul network, which carries traffic from the exchange to the core network and which is shared by all concurrent users. Therefore, in our opinion, and looking solely at telephone networks, the key element which differentiates current from next generation access networks is the type of infrastructure used. A NGA will in all likelihood depart from the current copper-based access network by making use of other technologies, such as fibre or wireless technologies.

The deployment of fibre in the access network may be one of the following: Fibre to the Cabinet (FTTCab) or Fibre to the Home (FTTH)[5]. As we can see from Figure 1, with FTTCab subscribers continue to be connected through their copper loops, but fibre is deployed all the way to the street cabinet (partly replacing the copper loop), i.e. much closer to the subscriber. Such an architecture allows an operator to deploy DSLAMs at that location. Street cabinets increase their importance significantly in the access network,

Figure 1. NGA architectures



in particular because they would then have active electronics in them. At the local exchange level, the Main Distribution Frame (MDF) may become superfluous, as the Sub-loop Distribution Frames (SDF) at the street cabinet level can take on their role in the network.

FTTH implies the complete replacement of the copper loop by fibre and SDFs and MDFs are no longer necessary. Under this architecture, there are two possibilities. In a point-to-point architecture, each subscriber has a dedicated fibre connection. Such an architecture allows for virtually unlimited traffic speeds (OFCOM (2007)). In a point-to-multipoint architecture (also known as a Passive Optical Networks (PON)[6]), a single fibre serves multiple subscribers (e.g. 32 or 64). In that case, that fibre would be split (using a passive optical splitter) at some point between the exchange and the customer's premises. This type of architecture would allow speeds of around 80Mbps per subscriber (with a 32 subscribers' split) (OFCOM (2007)). The current access network sites, such as the local exchange and the street cabinets, would still be useful. The local exchanges would typically contain the Optical Distribution Frames (ODF) which terminate the fibre loops and the street cabinets could contain the passive optical splitters necessary for a point-to-multipoint FTTH architecture.

If we look at the services which a NGA is likely to provide, there are already other networks capable of providing them, in particular, cable networks. The architecture of cable networks, typically hybrid fibre-coaxial (HFC) networks, is similar to an FTTCab deployment: fibre is deployed to a street cabinet level and coaxial cable is used to connect to each subscriber[7]. The current technology used is DOCSIS 2.0 and allows for speeds up to 50Mbps downstream and 30Mbps upstream, although the next version (DOCSIS 3.0) allows up to 220Mbps downstream and 120Mbps upstream (BSG (2007)). However, DOCSIS 3.0 entails investment in new head-end equipment and upgrades to customer premises equipment.

Our view is that a NGA will be more similar to the current cable architecture and, as such, a NGA based on cable is likely to be only a small upgrade from the current deployment, a view also shared by OFCOM (2007). This implies that the investment cable networks need to make to deploy a NGA is much smaller than that of copper-based access network owners (BSG (2007)).

Wireless alternatives already exist (e.g. WiMAX[8] and WiBro[9]) to reach subscriber's homes. However, there is some degree of scepticism as to whether these are viable alternatives to wired broadband. BSG (2007), OFCOM (2006a) and OFCOM (2007) mention various problems such as coverage, contention, symmetry and speed which may not allow for triple play services to be provided by such wireless technologies. Additionally, and perhaps most importantly, spectrum availability may be an important and possibly insurmountable constraint. Therefore, in our opinion, such technologies are more likely to be a complement to a NGA rather than a substitute.

NGAS IN CONTEXT

Value Chain and Firm's Strategic Behaviour

NGAs allow multiple services to be offered to consumers under different industry arrangements. The boundaries of a NGA operator are not clear yet.

In short, NGAs allow functions to be separated in two different hierarchically ordered layers: the basic and media layers, comprising the transmission infrastructure; and the contents layer, encompassing communication and media services that can offered on such an infrastructure. Contrary to the PSTN architecture, where the network is tightly coupled with the (telephone) services it provides, a NGA is inherently a flexible platform for service delivery.

In a NGA, the service-related functions are independent from the underlying transport-related technologies, and this allows the uncoupling of service provision from transport. Thus, it is possible that users have access to a wide range of services from different service providers.

The NGA value chain can be broken down horizontally and vertically. On the horizontal line, there are essentially three elements: customer equipment, network equipment, and interconnection with other networks. Vertically, as aforementioned, there are two layers: the service and the transport layer.

It is on the vertical layer that most changes are expected to occur. The innovative separation of infrastructure from service allows firms to undertake new strategies, which means that the boundaries of the firm will be redesigned, as competitors fit into the value chain. The transport layer is no longer coupled with the service provided. The current division of the transport layer into access and core is likely to be challenged as the services provided do not impose the same constraints as under the PSTN architecture.

Today, because most services are tightly coupled with the transport network, regulation has been applied to the bundle transport-services. NGAs allow regulation to be more service-neutral. Users are likely to combine different service providers and, as such, NGAs may have to support various contractual arrangements to meet consumers' dynamic needs. In turn, this opens up opportunities for market players to operate in specific network elements and services.

Broadly speaking, operators are likely to deploy NGAs where the gain from doing so outweighs the costs. In that context, there are several interesting questions which have to be raised and which are likely to influence NGA deployment decisions.

Firstly, NGAs are costly and are likely to be deployed in densely populated areas or for new build premises, as the incremental cost is relatively small in those situations compared to current access technologies (OFCOM (2007)). For existing (or potential) customers, operators have two available strategies (and it is not clear yet which they will choose): NGAs could be rolled out in areas where customers already have the best possible access technologies, i.e. areas with short loops where ADSL already delivers high speeds; or NGAs could complement existing access technologies by delivering higher speeds for customers who currently have longer loops and cannot experience the best ADSL products.

Secondly, NGA roll-out is more likely in countries where cable exerts a higher competitive pressure, i.e. telephone operators are likely to deploy NGAs in countries or regions where cable operators are trialling network upgrades which will allow them, in the future, to provide equivalent services to those of an NGA operator[10]; additionally, such a roll-out is also more likely where there is more to gain by selling additional (and currently unavailable) services such as IPTV[11] (OFCOM (2007)).

Thirdly, wholesale revenues are likely to see their importance increased for vertically integrated operators: the expected intensive downstream competition gives them an incentive to look more for wholesale revenues and thus try to recoup the investment made in NGAs (BSG (2007)).

Fourthly, in some countries, the motivation for NGA deployment may be the gradual evolution of the network. For instance, Marcus and Elixmann (2008) argue that the deployment of a NGA in the Netherlands using FTTCab is the appropriate business strategy, although such an NGA will provide ADSL2+-type services (in terms of speed) because sub-loop lengths are still high.

Fifthly, Amendola and Pupillo (2008), basing their assessment on JP Morgan (2006), put forward the idea that incumbents' strategies regarding NGA deployment may be materially different from new entrants' strategies. In particular, incumbents may prefer to choose FTTCab whereas new entrants may choose FTTH, bypassing the incumbent's network. For incumbents, FTTCab allows the provision of improved or additional services which are likely to raise more revenue, at the same time as it may allow cost savings in operating expenditure: this a "gradual evolution" strategy which does not allow the provision of FTTH-like speeds and services, but which also does not imply significant investments. By contrast, provided new entrants have access to ducts in metropolitan areas – the main cost driver in NGA deployment –, as well as a sufficient market share, there may be a business case for FTTH deployment[12].

Vertical Integration and Bottlenecks

The advantages of vertically integrated NGAs are not clear cut. Operators may adopt a comprehensive approach – encompassing both the infrastructure and service layers – or there could be specialization - different operators for the infrastructure and service layers.

Theoretically, there may be a rationale for the regulator to interfere in the industry where there are bottlenecks. Considering that regulation is typically a second-best solution, the regulator should consider what incentives should be in place so as to permit innovation in both the infrastructure and services layers.

In telecommunications, a bottleneck exists when it is not technically or economically feasible to duplicate some infrastructure elements. In the NGA case, the transport layer is at the core of the debate.

The European Commission (2003) Recommendation identifies barriers to entry into a market (or the possibility of overcoming such barriers within a certain time period) as a key criteria to decide whether a market should or should not be susceptible to ex ante regulation, i.e. whether the market is an economic bottleneck. The existence of economies of scale, economies of scope and high sunk costs are commonly associated with barriers to entry. But, as the European Commission (2003) notes, barriers to entry may also be of a legal or regulatory nature. As noted by Cave (2004), these types of barrier to entry can be non-strategic, i.e. they are not created by operators themselves.

Economies of scale are usually associated with fixed costs. But as Schmalensee (2004) observes, economies of scale are not sufficient to allow incumbents to raise prices above cost without attracting entry; it is only when those costs are sunk[13] that such a behaviour is to be expected.

Undoubtedly, a NGA requires a substantial upfront investment by operators (BSG (2007), Williamson (2007)). Whether the transport layer of a NGA is likely to be a bottleneck depends on the "sunkness" of that investment. In much the same way as the traditional access network was regarded has a highly sunk investment (e.g. Pindyck (2007)), our view is that the investment in a NGA is largely sunk. As such, the transport layer of a NGA is likely to be an economic bottleneck, a point also made by the ERG (2007).

On the issue of vertical integration, there is a concern that the operator controlling the bottleneck might enjoy a competitive advantage. If there are complementarities between the infrastructure and service layers (e.g. demand complementarities or economies of scope), vertical integration may be a sensible strategy and, as such, incumbents might want to keep any future developments in the infrastructure to themselves, claiming this is necessary to incentivise investment and innovation. On the other hand, new entrants view this development as a strategy for incumbents to reinforce their market power. As we have previously argued, a NGA will effectively allow a more clear-cut separation between the infrastructure and service layers, which suggests that these two layers will enjoy fewer complementarities than they did thus far.

This leads us to conclude that the concept of bottleneck in NGAs is likely to be dynamic. If the transport and service layers become more separate, telecom operators are likely to develop independent strategies for each of these. For example, contrary to the PSTN, full control of a given communication network to provide the service to final consumers might not be needed or even desirable. Indeed, as interconnection on a NGA is facilitated, it is quite likely that, under certain circumstances, multiple alternative platforms and protocols exist, which allow an operator to choose the most suitable connection for a given client.

Thus, the technological innovation and the interconnection development might sweep aside the traditional response to bottlenecks. This layer separation challenges the current regulatory environment, as it allows regulators to be more service-neutral, focusing instead on guaranteeing access to the infrastructure while allowing competition on the service level.

The European Regulatory Framework

The European Regulatory Framework (ERF)[14] regulates electronic communication services in Europe. The rationale of the regulatory approach is to make telecommunications markets more competitive, so that as dominant operators see their market power reduced, reliance on competition rules will increase while sector-specific regulation is gradually phased out.

Regulatory obligations are imposed in three steps: firstly, markets susceptible to ex ante regulation are defined; secondly, "significant market power" (SMP) operators are identified in those markets; and finally, regulatory obligations are imposed on the SMP operators.

Markets susceptible to ex ante regulation are those where there are significant barriers to entry which are unlikely to be overcome within a relevant time horizon and where competition law alone is unlikely to solve the problem. In this context, the European Commission (2003) adopted a Recommendation (amended by the European Commission (2007b) Recommendation) that defined the product and service markets susceptible to ex ante regulation.

The regulatory obligations, which can be imposed on SMP operators are: transparency; non-discrimination, accounting separation, compulsory access, and cost orientation as well as any other obligations in exceptional circumstances. Under its amendment to the ERF (European Commission (2007a)), functional separation may also be imposed.

The underlying rationale of the ERF is based on the ladder of investment principle (see, for example, Cave (2006) or ERG (2007)). The regulatory regime is designed to encourage investment by both incumbents and new entrants. As for the latter, the significant necessary investments to provide telecommunications services are unlikely to be made on a one-off basis. What is likely to happen is that new entrants enter the market slowly and invest in capital assets gradually, as their customer base also increases. But in order to do so they need to be able to provide services at a time where their networks are not yet fully built and hence may require several complementary services from incumbents. As new entrants climb the "ladder of investment", they slowly reach final customers with their own networks, competing directly and on several dimensions (e.g. price, quality, coverage, etc) with incumbents. In effect, as the ERG (2004) points out, "... such interventions have as their long-term objective the emergence of self-sustaining effective competition and the ultimate withdrawal of regulatory obligations" (p. 68). In essence, the ultimate aim of the regulatory framework is to promote network competition.[15]

Until such effective network competition materialises, the ERF is a "substitute". As such, it tries to emulate what would happen under true network competition. The underlying idea of promoting competition at the deepest possible level in the network is an attempt to give new entrants as much flexibility as possible in the design (characteristics, price, quality, etc) of the services they sell to final customers.

Wholesale Markets

Market definition, which has both a product and geographic dimension, starts from end-user demands (retail markets); network or wholesale inputs are a derived demand. Hence, if two services provided over different platforms are in the same market, then inputs into those services (e.g. wholesale products) would also belong to the same market.

Broadly speaking, there are three types of wholesale services which are likely to be provided over a NGA: existing services, better quality versions of existing services and new services (OFCOM (2006a)). There is, however, considerable uncertainty as to what these services may actually be in the future. For existing services, it could be argued that although the technology used to deliver them over a NGA is different, the services themselves are the same and, as such, will belong to the same market as they do today. Better quality services may belong to the same market as current wholesale services if they are substitutable (price constrained).

Finally, new services unique to NGAs may belong to a separate market. For instance, Cave et al (2006) note that a fibre-based network will not only allow the provision of the same services provided today, but also of new services such as IPTV. In that case, a separate wholesale market would also need to be defined, which could be considered as "emerging" and hence not subject to ex ante regulation. But this input may also be used as an input for other retail services (such as those mentioned previously), which raises the question of whether the new services are sufficiently important for this to be a separate market.

A NGA will affect essentially two markets contained in the European Commission (2003) Recommendation: the market for wholesale unbundled access (including shared access) to metallic loops and sub-loops (market 11) and the market for wholesale broadband access (market 12). The European Commission (2007b) Recommendation replaces the 2003 Recommendation and maintains these two markets as markets susceptible to ex ante regulation. However, and following the arguments put forward by the ERG (2007) and Cave et al (2006) (among others), the first market is now the market for wholesale (physical) network infrastructure access (including shared or fully unbundled access) at a fixed location. The second market (wholesale broadband access) maintains its designation.

ERG (2007) argued that the European Commission (2003) definition of market 11 excluded fibre access by focusing solely on metallic loops and would thus have to be enlarged. Such a definition should be sufficiently broad so as to include the most likely network layouts under discussion: FTTCab (where local loops would still be partly copper-based) and FTTH (where loops would be completely fibre-based). As OFCOM (2006b) noted when conducting its wholesale local access review, although copper and cable local access services belonged to the same market (i.e. fibre access was excluded), that does not mean that a NGA based on fibre would not be subject to ex ante regulation – after deployment, a new market review needs to be carried out so as to analyse the inclusion/exclusion of fibre access in the relevant local access market.

As for market 12, the ERG (2007) argued that the European Commission (2003) definition included the provision of wholesale access services between the customer's premises and the DSLAM or higher levels of the IP network and this would be broadly maintained under a FTTCab or FTTH architecture. As such, the deployment of NGAs would not affect the definition of this market.

Geographic market definition, however, is very likely to change significantly with the emergence of NGAs. Geographic markets are defined where there is some homogeneity in the competitive conditions within that market. Therefore, several neighbouring geographic markets may exist, in which case the competitive conditions must be similar within the market but clearly different across markets (see, for example, European Commission (1997, 2002)).

Competitive conditions refer, among other things, to possible differences in the costs of providing the service or existing or potential competition (namely the presence of cable providers) (Amendola and Pupillo (2008) and BSG (2007)). On the cost side, it is likely that there are joint and/or common costs to serving neighbourhoods or even cities[16]. On the competition side, it is quite possible that some geographic markets will experience strong network competition due to the presence of cable providers whilst others may have less or no network competition at all (BSG (2007)).

Regulators are well aware of this. The key aspect of OFCOM's (2006b) review was the introduction of geographic wholesale markets, based on competitive local conditions, an approach also followed by CMT (2007) (the Spanish regulator). In particular, OFCOM (2006b) concluded that internet provision through cable networks was in the same retail and wholesale market as ADSL broadband internet access – even though cable and ADSL broadband are provided over different networks, they exert sufficient competitive pressure on each other to belong to the same product market. From a geographic point of view, OFCOM (2006b) sought to identify markets where competitive conditions were relatively homogeneous: the main competitive factors identified were the availability of cable-based services, the current or planned availability of local loop unbundling (LLU)-based services or the likelihood of LLU-based entry, as well as the existence of common pricing constraints. Market power was then assessed in each of these geographic markets and the imposition (or not) of regulatory remedies could well differ between markets as pointed out by Amendola and Pupillo (2008), BSG (2007) and Waverman and Dasgupta (2006).

REGULATORY CHALLENGES OF NGAS

Regulatory Approaches to NGAs

OFCOM (2006a) reiterates that the promotion of competition at the deepest possible level in the network is expected to lead to a "ladder of investment". If the same principle is applied to NGAs, then there would be various levels where competition may materialise:

- **Ducts and trenches:** The ability to use the incumbent's ducts and trenches to roll out a NGA significantly reduces deployment costs and allows operators a choice of different technological solutions for service provision;
- **Sub-loop unbundling:** This would be equivalent to local loop unbundling, but would occur at the cabinet (closer to the customer); this option has various problems, because it requires other operators to invest significantly in order to move closer to the customer, thus requiring a higher market share to compensate for such an investment; in addition, there may be space constraints at the cabinet level;
- **Gibre unbundling at the exchange:** This is only possible with point-to-point fibre; in addition, operators may find it worth investing only if part of that investment is offset by the sale of exchanges (which would no longer be necessary);
- **Wholesale bitstream products:** If competition deeper in the network is not expected to deliver benefits, offering wholesale interconnection products may be a good alternative, although at the cost of less innovation in NGA services.

According to the ERG (2007), the possibility of unbundling at the physical layer depends crucially on the NGA layout. With FTTCab, unbundling of local loops would be possible in the same way as it is today, but closer to the customer – in street cabinets. However, Marcus and Elixmann (2008) argue that with FTTCab the standard regulatory remedies are harder to apply because of the geographic dispersion of street cabinets (which become the points of interconnection) as well as the likely restrictions in terms of access and space these have. Even when it is possible to unbundle at the street cabinet level, there is a problem of market size. The first mover – i.e. the operator that sets up FTTCab first – might have an advantage if the market is not large enough to accommodate more than one operator and competition would thus be compromised.

In addition, and according to the ERG (2007), it is important to consider the need for a backhaul between the street cabinet and other operators' networks, which raises the possibility of implementing duct sharing. The change in infrastructure may affect LLU, because MDFs will be reconfigured or even phased-out. There must be a balance between the freedom of operators to roll-out NGAs and the need for promoting competition. The worry is that NGAs have significant economies of scale, economies of scope with incumbents' facilities and few access alternatives (e.g. wireless)[17].

For FTTH deployment, construction costs make up the majority of the investment necessary. As such, duct sharing could be imposed on a symmetric basis on all operators. With FTTH, the unbundling possibilities depend on the architecture chosen: if a point-to-point solution is adopted, it is possible to unbundle fibre loops in much the same way as copper loops; if a point-to-multipoint solution is adopted, then a natural unbundling frontier would be at the passive optical splitter level (ERG (2007)).

Following OFCOM (2006a), the regulatory mechanisms commonly used are:

- Mandatory and non-discriminatory access but no price regulation
- Retail-minus prices
- Cost-based prices
- Cost-based prices with activity-specific cost of capital
- Adjusted returns

Imposing mandatory and non-discriminatory access but abstaining from regulating prices would allow operators to decide whether to invest on NGA depending on their expected returns from the investment. Such an approach, however, tilts the balance in favour of vertically-integrated incumbents who can still distort competition in their favour, either through high prices for access, cross-subsidies to the downstream division or margin squeeze competitors (OFCOM (2006a)). Such a strategy by the vertically integrated operator would have the objective of foreclosing the retail market and result in a monopoly situation.

Retail-minus (wholesale) prices are based on retail prices less the necessary retail margin for other operators to supply the service profitably. Such a pricing mechanism is normally used when (wholesale) competition is expected to emerge in the future, thus driving wholesale prices towards costs, or for new and innovative markets. Provided there is no margin squeeze, retail-minus prices is likely to result in efficient investment, as there is no cap to investor returns.

Cost-based prices are usually calculated using Long-Run Incremental Cost (LRIC) principles, and thus reflect the replacement cost of the assets used by a given service, including a (company-level) cost of capital which reflects the associated risk. By contrast, in specific cases where there is a higher or lower systematic risk (which cannot be diversified), an activity-specific cost of capital may be used.

When prices are regulated, the investor faces asymmetric returns as he is allowed to earn a predetermined (regulated) return if the investment is successful, but has to support all the losses if it is not. This results in reduced incentives to invest which may be corrected through prices which adjust returns for this asymmetry (OFCOM (2006a))[18].

The migration from existing regulation must consider the investment made by competitors, whether that investment was recouped and whether such competitors will continue to have access to wholesale products equivalent or better than the current ones. A balance between costs savings coming from the NGA and eventual risks to competition must also be struck.

As mentioned earlier, a NGA will continue to be an input to existing services as well as a means to deliver entirely new services. In terms of regulatory remedies, this raises the possibility of mandated access but only up to certain speeds (the existing services) (Cave et al (2006)). Williamson (2007) puts forward the argument that anchor product regulation – the creation of equivalent products to those existing today, but delivered through the NGA - is a reasonable regulatory remedy for the problems raised by a NGA. As proposed by Williamson (2007) and OFCOM (2007), anchor product regulation is a mix of the two extreme regulatory solutions: price regulation and forbearance. On the one hand, it protects consumers and other operators by requiring the offer of equivalent services through the NGA, possibly at regulated prices, to those existing today; on the other hand, it leaves the NGA owner with sufficient flexibility regarding the price and product characteristics for the new services a NGA will allow. Needless to say, the existence of anchor products would in itself constrain the pricing of the new services if they are in some way substitutable.

Blum et al (2007) show that the regulatory threat of intervention may be sufficient to prevent an operator from charging too high a price. This is particularly relevant in the light of the German case of allowing a regulatory holiday on operators deploying NGAs. The regulator can influence both the likelihood of NGA investment as well as their respective prices even without imposing price regulation.

Interestingly, NGAs make forbearance a more viable alternative (provided, of course, access is provided on a non-discriminatory and equivalent basis to all third parties), as the NGA owner will want to recoup its investment and hence carry as much traffic as possible, even if such traffic belongs to rival operators (OFCOM (2007) and BSG (2007)).

Partly, the regulatory problem presented by NGAs verges on what is likely to happen regarding the voluntary provision of wholesale access products. If the NGA owner decides to create a wholesale product so as to recoup its investment, then there may be no need for regulatory intervention. The issue is similar to the incentives faced by mobile operators to allow the creation of mobile virtual network operators (MVNOs): in that case, the creation of the wholesale product was almost never a natural market outcome and had to be mandated. The case of cable is similar: such a wholesale product does not exist (although no mandated access has been determined in virtually all countries).

We conjecture that initially the NGA owner will be reluctant to create a wholesale product which would allow competitors to access potential retail customers. However, after some time has elapsed and the NGA owner has already gained a sufficient customer base, it is unlikely that the current offer (product characteristics, prices, etc) will attract further customers. This is normally the phase where product differentiation/market segmentation becomes attractive. The NGA owner can himself create such differentiated products to increase penetration (much in the same way as mobile operators did by creating “secondary brands” using the same network) or create a wholesale product and allow other operators to carry out that task. The underlying assumption is that downstream variety increases upstream demand and this variety is better provided by multiple firms[19].

In turn, if this line of thought materialises, a regulatory holiday such as that suggested in

Germany may be the best regulatory strategy. After that regulatory holiday, it may not even be necessary for the regulator to mandate access (or to announce at the outset that it will mandate access after the regulatory holiday): regulatory threat has worked before (e.g. on BT in the UK) and the incumbent created such a product even in the absence of "actual" regulation. This suggests that regulators should, at this point, provide clear policy indications of what they are likely to do (and why) once the NGA has been deployed.

Is Moving up the Ladder Necessarily Bad?

Even if a NGA is a bottleneck, this does not necessarily imply the existence of less competition, but rather of a different type of competition (Williamson (2007)). For instance, switching service providers may become easier, faults associated with local loop unbundling may become less likely and entry may occur in the markets for services and applications which a NGA will allow.

The design of the NGA described by Openreach (2007a, 2007b) suggests that this may indeed be so. Openreach (2007a) is consulting on the deployment of FTTH using a point-to-multipoint architecture. The equipment at the customer's premises would support up to two service providers, one for voice and one for data[20].

Thus, and although the ladder of investment principle is still appropriate to deal with enduring economic bottlenecks, with NGAs we are likely to see a move up the ladder, i.e. the deepest level in the network where competition will be promoted may not be as "deep" as it is today. In particular, we may see a move away from promoting competition at the physical layer towards competition in upper (service) layers (e.g. moving away from ULL towards wholesale broadband access). This is particularly so if MDFs are gradually phased-out. In that case, a stronger focus on wholesale broadband access is likely to materialise, as regulators will push for a design of such products which emulates as much as possible the choices operators could make if competition could take place at a deeper (e.g. physical) level in the network.

In fact, OFCOM (2007) suggests that NGA deployment is likely to shift regulation from passive input regulation (such as duct, copper lines or optical fibre) to active input regulation (equipment such as DSLAMs or Optical Line Terminals), as the viability and effectiveness of passive input regulation is likely to be reduced. OFCOM (2007) discusses the limitations of passive input competition as network deployment moves away from densely populated areas. In such areas, it may be more appealing to use active input products. Provided such active input products are configurable in as close a way as possible to what another operator could achieve with passive inputs, the benefits of competition are likely to emerge without the associated costs of passive input use[21].

The Investment/Competition Dilemma

OFCOM (2006a) suggests that if NGAs are a bottleneck, then regulators today would face a serious problem. The usual remedies for such markets, such as mandated access and regulated prices (cost-based prices, retail-minus or any other approach), would reduce investor returns. NGAs cause a regulatory dilemma (Laffont and Tirole (2001)): regulation designed to promote competition after network deployment (such as mandated access and regulated prices) will reduce the incentives to invest in the first place; if the regulatory approach wishes to promote investment, then this is likely to be coupled with reduced competition after network deployment. This problem makes it important to clarify the regulatory approach to be followed after network deployment before the investment takes place.

Gayle and Weisman (2007) discuss the trade-offs between static and dynamic efficiency. In particular, an analogy is made with policies regarding copyrights, patents and trademarks, where distortions to static efficiency are accepted (e.g. prices above marginal costs) so as to encourage dynamic efficiency through investment in innovation. Gayle and Weisman (2007) argue that when unbundled network elements are price regulated, incumbent have less incentives to invest and hence dynamic efficiency is sacrificed. Jung et al (2008) find some empirical support for this claim. Pindyck (2007) argues that mandatory unbundling at cost-based prices discourages investment by incumbents and new entrants, as the adoption of current cost principles in a market which is permanently faced with a decreasing price trend does not allow the investor to recoup its investment. As Hausman (1997) points out, the dynamic effects of regulation are often much larger than the static effects of monopoly power and hence the regulator should err on the side of leniency to encourage new products[22].

NGA roll-out is expected to have significant risks. Under regulation, this risk is likely to be shared between the operators and consumers: consumers would support all the risk if operators could freely define prices and operators would support most of the risk if prices were regulatory constrained to reflect costs. Here we have a dilemma: if operators were free to set prices, one would expect a relatively quick NGA roll-out, which would deliver dynamic efficiency at the cost of allocative efficiency (operators would charge prices in excess of costs and extract consumer surplus). By contrast, a cost-based price is likely to reduce incentives for NGA deployment, as the returns to investment are capped above but not below (OFCOM (2007)). Even if only mandated access to NGAs is imposed, Waverman and Dasgupta (2006) stress that this may reduce both incumbents' and new entrants' incentives to invest in their deployment.

On this relationship between investment and the regulatory regime, BSG (2007) notes that where network competition is considered to be sufficiently intense, such as in the USA, Canada and Hong Kong, regulators have adopted a liberal approach and not required network owners to provide access to other operators. Naturally, such an approach has a positive effect on investment incentives. In particular, in Canada, a Regulatory Review Panel concluded that unbundling based on the ladder of investment principle could undermine the achievement of genuine network competition. In the Netherlands and Belgium, where cable competition is strong, regulators are considering whether to impose regulatory obligations on operators deploying NGAs.

The US regulatory approach is based on permanent forbearance, i.e. there is no regulatory requirement for the network owner to provide access to third parties. Such an approach has the great merit of delivering the right investment incentives, making the investor fully liable to potential losses but also allowing him to retain any gains obtained. Forbearance could also be temporary, where the regulator explicitly abstains from regulating services for a given period of time (e.g. the case of Germany for the roll-out of NGAs). This would be similar to the patent regime. The success of forbearance as a regulatory approach depends on the likelihood of network competition: if network competition is unlikely to materialise, then any type of forbearance will distort the market: prices will be higher than socially optimal (allocative inefficiency) and there may be over-investment. If network competition is expected to emerge after a given period of time, then temporary forbearance may give the NGA owner sufficient assurances to invest in the first place.

CONCLUSION

The introduction of NGAs is likely to reshape the telecommunications sector. Whilst the NGA infrastructure is likely to be a bottleneck and, thus, attract regulatory scrutiny, service operation is likely to be more infrastructure-independent and, as such, raise questions as to the desirability of the vertically integrated incumbent operator model.

There is, however, substantial uncertainty as to how the market will evolve. This suggests that regulators should thread carefully. In particular, the ladder of investment principle is likely to be affected by NGAs, but this is not necessarily bad from a competition point of view. Moreover, the fact that NGAs are yet to be deployed raises the investment/competition dilemma: if regulators impose (or threaten to impose) regulatory remedies after NGA roll-out, this will obviously affect incentives to deploy such networks in the first place. Whilst it is desirable to promote competition in any sector (static efficiency), substantial benefits often accrue from investment in innovations (dynamic efficiency). This trade-off is at the heart of the discussion of the regulatory challenges raised by NGAs.

REFERENCES

Amendola, G. B., & Pupillo, L. M. (2008). The Economics of Next Generation Access Networks and Regulatory Governance: Towards Geographic Patterns of Regulation. *Communications & Strategies*, 69, 85-105.

Blum, U., Growitsch, C., & Krap, N. (2007). Broadband Investment and the Threat of Regulation: Preventing Monopoly Exploitation or Infrastructure Construction? *Review of Network Economics*, *6*(3), 342-354.

Broadband Stakeholder Group (BSG). (2007). Pipe Dreams? *Prospects for next generation broadband deployment in the UK*. London: BSG.

Cave, M. (2006). Encouraging network competition via the ladder of investment, *Telecommunications Policy*, *30*, 223-237.

Cave, M., Stumpf, U., & Valletti, T. (2006). *A Review of certain markets included in the Commission's Recommendation on Relevant Markets subject to ex ante Regulation*. Report for the European Commission.

Cave, M. (2004). Economic Aspects of the New Regulatory Regime for Electronic Communications Services. In P. Buiges & P. Rey (Eds), *The Economics of Antitrust and Regulation in Telecommunications* (pp. 27-41). E. Elgar.

Comisión del Mercado de las Telecomunicaciones (CMT). (2007). Consulta publica sobre redes de acceso de nueva generacion.

European Regulators Group (ERG). (2007). ERG Consultation Document on Regulatory Principles of NGA, ERG (07) 16.

European Regulators Group (ERG). (2004). ERG common position on the approach to appropriate remedies in the New Regulatory Framework. *ERG 30*(03).

European Commission. (2007a). Proposal for a directive of the European Parliament and of the Council amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and services, and 2002/20/EC on the authorisation of electronic communications networks and services.

European Commission (2007b). Recommendation 2007/879/EC of 17 December 2007 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services. *OJEU*, L344/65.

European Commission (2003). Recommendation 2003/311 of 11 February 2003 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services. *OJEC*, L 114/45.

European Commission (2002). Guidelines 2002/C 165/03 on market analysis and the assessment of significant market power under the Community regulatory framework for electronic communications networks and services. *OJEC*, C165/31.

European Commission (1997). Notice on the definition of the relevant market for the purposes of Community competition law. *OJEC*, C 372.

Gayle, P., & Weisman, D. (2007). Efficiency Trade-Offs in the Design of Competition Policy for the Telecommunications Industry. *Review of Network Economics*, *6*(3), 321-341.

Hausman, J. (1997). Valuing the effect of regulation on new services in telecommunications. *Brookings Papers on Economic Activity, Microeconomics*, *28*(1), 1-54.

JP Morgan (2006). *The fibre battle*.

Jung, I., Gayle, P., & Lehman, D. (2008). Competition and investment in telecommunications. *Applied Economics*, *40*, 303-313.

Laffont, J. J., & Tirole, J. (2001). *Competition in Telecommunications*. MIT Press.

Marcus, J. S., & Elixmann, D. (2008). Regulatory Approaches to Next Generation Networks (NGNs): An International Comparison. *Communications & Strategies*, 69, 19-40.

OFCOM (2007). Future broadband policy approach to next generation access. Public consultation document.

OFCOM (2006a). Regulatory challenges posed by next generation access networks. Public discussion document.

OFCOM (2006b). Review of the wholesale broadband access markets 2006/07. Public consultation document.

Openreach (2007a). Openreach services delivered over Fibre To The Premises (FTTP). Consultation document.

Openreach (2007b). Product definition overview: Generic Ethernet Access delivered over Fibre to the Premises.

Pindyck, R. (2007). Mandatory Unbundling and Irreversible Investment in Telecom Networks. *Review of Network Economics*, *6*(3), 274-298.

Schmalensee, R. (2004). *Sunk Costs and Antitrust Barriers to Entry*. American Economic Review, 94(2), 471-475.

Waverman, L., & Dasgupta, K. (2006). Investment in telecommunication networks: A 21st Century perspective. Unpublished manuscript.

Williamson, B. (2007, March). *Risk, reward and efficient investment in access networks*. Paper presented at an OFCOM European seminar on regulatory challenges posed by next generation access networks, London, United Kindgom.

KEY TERMS

ADSL: ADSL (Asymmetric Digital Subscriber Line) allows voice and data to be sent simultaneously over the existing copper-based telephone line (ITU-T Recommendation G.992.1), with speeds up to 6Mbps downstream and 640Kbps upstream. In July 2002, the ITU completed G.992.3 and G.992.4, two new standards for ADSL technology collectively called “ADSL2”, allowing for a maximum 12Mbps downstream speed and 1.3Mbps upstream. In January 2003, a new standard called ADSL2+ (ITU G.992.5) doubled the bandwidth used for downstream data transmission, achieving rates of 20 Mbps on phone lines as long at 1,5Km.

FTTCab: FTTCab (Fibre-to-the-Cabinet) is a network architecture where optical fibre replaces copper in the access network connection between the local exchange and street cabinets.

FTTH: FTTH (Fibre-to-the-Home) is a network architecture which makes use of optical fibre to connect to the subscriber's homes.

IPTV: IPTV (Internet Protocol Television) is a digital television service delivered through a broadband connection.

Ladder of Investment Principle: Regulatory principle with the objective of encouraging investment by both incumbents and new entrants and ultimately promoting network competition. It is based on the idea that investments by new entrants are gradual as their customer base increases, but until this happens several complementary services (provided by incumbents) are needed.

Next Generation Access Network (NGA): Not fully copper-based access network, capable of providing broadband access services with sustained bandwidths clearly higher than those available with fully copper-based access networks.

Next Generation Network (NGN): A multi-service (voice, data, video, etc) packet-based network in which service-related functions are independent from the underlying transport-related technologies.

NGA Layers: The basic and media layers comprise the transmission infrastructure and the contents layer encompasses the communication and media services that can offered on such an infrastructure.

Triple Play: Provision of three communications services (voice, data and television) over a single broadband connection.

VDSL: VDSL (very high speed digital subscriber line) download speeds are in the 13-55 Mbps range, over short distances, usually between 300 and 1500 meters, using twisted pair copper wire. The shorter is the distance, the faster the connection rate. VDSL2 (ITU-T G.993.2) allows speeds up to 100Mbps each way, but depends crucially on the length and quality of the copper loop. For loops longer than approximately 1.6Km it no longer provides a faster bit rate than ADSL2+.

ENDNOTES

1 HDSL (high-data-rate digital subscriber line) delivers 1.544 Mbps of bandwidth each way over two copper twisted pairs. The operating range of HDSL is limited to approximately 3,5Km.

2 SDSL (single-line digital subscriber line) delivers 1.544 Mbps both downstream and upstream over a single copper twisted pair. The use of a single twisted pair limits the operating range of SDSL to approximately 3Km. This type of service uses all the bandwidth available and, therefore, does not allow the telephone service to be provided simultaneously.

3 High-definition television (HDTV) is a digital television broadcasting system with a greater screen resolution than traditional systems.

4 For instance, in the UK only about 50% of households will be able to access 8Mbps or more (BSG (2007)).

5 There are other possibilities, such as Fibre to the Building (FTTB); the key difference in all variants is the end-point of the fibre (cabinet, building, home).

6 Different PON systems exist: Broadband PON (BPON), Gigabit PON (GPON) and Ethernet PON (EPON). EPON makes use of an Ethernet router at the street level instead of a passive optical splitter.

7 Contention is a well known problem for cable networks as subscribers have to share the available capacity at the street cabinet level.

8 **W**orldwide **I**nteroperability for **M**icrowave **Ac**ce**s**s (WiMAX) is based on the IEEE 802.16 standard. According to the WiMAX Forum, "WiMAX provides fixed , nomadic, portable and, soon, mobile wireless broadband connectivity without the need for direct line-of-sight with a base station. In a typical cell radius deployment of three to ten kilometers, WiMAX Forum Certified™ systems can be expected to deliver capacity of up to 40 Mbps per channel, for fixed and portable access applications."

9 WiBro (**Wi**reless **Bro**adband) is based on the same standard as WiMAX (802.16), but adapted to provide connectivity on the move (such as that required by mobile phone access).

10 The rationale for NGA deployment in such a scenario is the "first mover advantage" that telephone operators would enjoy.

11 IPTV (Internet Protocol Television) is a digital television service delivered through a broadband connection.

12 Amendola and Pupillo (2008) mention two such examples: Iliad in France and NetCologne in Germany.

13 Sunk costs are costs which cannot be avoided if a firm stops producing, i.e. they are not recoverable. They are usually associated with specific assets which have little or no alternative use in other activities.

14 The European Regulatory Framework includes Directive 2002/21/EC ("Framework Directive"), Directive 2002/20/EC ("Authorisation Directive"), Directive 2002/19/EC ("Access Directive"), Directive 2002/22/EC ("Universal Service Directive"), Directive 2002/58/EC ("Data Protection Directive") and Directive 2002/77/EC ("Competition Directive"). The European Commission is currently consulting on a proposal to amend the ERF (European Commission (2007a)).

15 It should be noted that this theory largely proved to be a failure in the United States, following the Telecommunications Act 1996: there is little evidence that the unbundling and resale regimes were used temporarily while new entrants built their networks. We thank a referee for pointing this out.

16 We thank a referee for pointing this out.

17 We thank a referee for this observation.

18 This type of pricing, suggested by OFCOM (2006a), could take the form of a glide path from higher/lower expected returns towards the expected returns. When returns are higher than expected, the glide path would allow the operator to keep a proportion of those returns for a longer period.

19 We thank a referee for this observation.

20 Openreach is not considering the deployment of copper-based services to such customers nor any variant of LLU or wholesale line rental.

21 OFCOM (2007) proposes the creation of such active line Ethernet-based access products both for FTTH and FTTCab deployments (although sub-loop unbundling, a passive input product, is also suggested).

22 We thank a referee for this observation.

Chapter V
The Impact of Government on the Evolving Market Structure of the U.S. Wireless Telephony Industry

Carol C. McDonough
University of Massachusetts Lowell, USA

ABSTRACT

The United States' wireless telephone industry has evolved from a minor segment of the communications industry to a major provider of voice, and increasingly data and video communication. The industry uses radiowaves to transmit signals, and radiowave spectrum is regulated by the federal government. Moreover, local transmission requires unobstructed antennae, which in rural and suburban areas has led to the construction of wireless towers. States and municipalities have sought to regulate the construction of such towers, citing issues of aesthetics and health. The development of the wireless industry has been constrained by such government regulation. This chapter discusses the impact of government on the market structure of the wireless industry.

INTRODUCTION

Wireless telephony has evolved from a minor component of the telephony industry to a ubiquitous component today. In the developing world, wireless communication provides access to telephony services without the cost of building out landline infrastructures. Mobile phone penetration rates, which measure the number of mobile phones per 100 population, are high in industrialized nations and lower in less developed areas. (Table 1). These numbers often reflect multiple phones per wireless subscriber.

Governments, at the federal, state, and local levels, have had a major impact on the structure of the wireless industry, through policies on licensing, mergers/acquisitions, pricing, and tower construction. This chapter discusses the impact

Table 1. 2007 Mobile phone penetration rates

Luxembourg	158%
Hong Kong	140%
Western Europe	110%
Europe total	95%
United States	85%
China	50%
India	23%
Worldwide	13%
Sources: BBC News,11/07; Informa,2007; Wikipedia; Fortune, 5/26/08	

of U.S. government policies on the structure of the U.S. wireless telephony industry.

EARLY HISTORY OF THE WIRELESS INDUSTRY

Because wireless telephony uses radio wavelengths for transmission, government regulation and distribution of wavelength affects competition in the industry. Internationally, the radio frequency (RF) spectrum is allocated by the International Telecommunication Union (ITU). Governments globally have taken on the assigning of frequencies, and this can be a highly political decision.[1]

Within the United States and its possessions, the RF spectrum is further allocated to non-Government and Government users. The Federal Communications Commission (FCC), acting under the authority of Congress, allocates frequencies to non-Government users. The National Telecommunications and Information Administration (NTIA) manages frequencies to departments and agencies of the U.S. Government**.** Table 2 details information on the allocation of RF spectrum in the United States.

In the United States in the 1970's, the FCC planned to license a single cellular system in each area, operated by the local wireline phone company. Experimental wireless licenses were granted in several areas. In some European countries, government-owned telephone companies, known as Post, Telephone, and Telegraph (PTT) companies, provided early wireless service.

In 1981, the FCC altered the cellular structure to a duopoly. 50 MHz of spectrum in the 800 MHz frequency band were set aside for two competing systems in each market, one for the local wireline telephone company and the other for a nonwireline company. In each of 734 cellular markets (306 Metropolitan Statistical Areas (MSA) and 428 Rural Service Areas (RSA)), the FCC used non-revenue generating processes such as comparative hearings and lotteries to select a nonwireline company. The FCC started to license cellular companies in 1982 and had essentially completed the licensing process by 1991.[2]

Similarly, the wireless industry in Canada, regulated by Industry Canada, expanded to two cellular telephone carriers in 1984. [3]

Another earlier form of mobile telephony is enhanced specialized mobile radio (ESMR), developed by assembling frequency bands originally allocated for two-way dispatch services. In the United States, ESMR operates in the 800MHz

Table 2. Allocation of radio spectrum in the United States

30 km	3 km	300 m	30 m	3 m	30 cm	3 cm	3 mm		
VLF	LF	MF	HF	VHF	UHF	SHF	EHF	Infrared	Visible Light
10 kHz	100 kHz	1 MHz	10 MHz	100 MHz	1 GHz	10 GHz	100 GHz		

range with 14 MHz of spectrum space. There usually is one ESMR service provider in a geographic area.

THE DEMAND FOR WIRELESS TELEPHONY INCREASES

By the 1990s, as technological advances introduced digital technology to the industry, the demand for wireless services had grown significantly. A developing literature advocated for the use of competitive bidding to assign wireless licenses. Early contributors to the literature on auctions as an efficient way to distribute resources in the public domain to promote social welfare included Herzel and Coase.[4] The argument for using the price system to allocate wireless spectrum was based on three types of efficiencies: elimination of rent dissipation associated with comparative hearings; assignment of licenses to the most productive suppliers, saving the transaction costs of secondary market redistributions; and the generation of revenues for public use.[5] Competitive bidding was adopted by New Zealand in 1989, India in 1991 and the United States in 1993. At least twenty-five other countries have instituted license auctions in recent years.

THE ROLE OF THE UNITED STATES' GOVERNMENT IN DETERMINING INDUSTRY STRUCTURE: PCS AUCTION STRUCTURE[6]

In 1993, Congress authorized the FCC to auction off broadband PCS licenses to operate in the 2 GHz band of the spectrum. PCS spectrum almost quadrupled the total amount of spectrum allocated to wireless service.

The FCC divided PCS spectrum into six frequency blocks (A-F), creating the potential for a maximum of six additional wireless providers per market. The wireless market was subdivided into 51 large blocks known as Major Trading Areas (MTA), and further subdivided into 493 smaller blocks known as Basic Trading Areas (BTA's). Blocks C and F were designated as entrepreneurial blocks, with upper limits set on bidders' gross revenues and total assets. Table 3 shows the PCS auction structure.

Each auction was conducted using the simultaneous multiple round auction process. Multiple round auctions provide more information to bidders about the value that other bidders place on the licenses. The likelihood that the licenses are acquired by those who value them most highly is increased, and greater revenue is gained. Simultaneous auctions are preferable to sequential auctions: simultaneous auctions increase the value of licenses by facilitating efficient aggregation and, by providing more information on the value of interdependent licenses, reduce the propensity of sophisticated bidders to bid cautiously to avoid the winner's curse[7]. Although the purpose of distributing PCS licenses via simultaneous multiple round auctions was to place the licenses in their most highly valued use, concerns arose that there were insufficient safeguards against reflexive bidding[8] and speculative bidding.

Once an auction had begun, generally one bidding round per day occurred. An auction closed if one round passed in which no bidder submitted a new acceptable bid on any license. Bidding stopped on all licenses simultaneously. [9]

Table 3. PCS auction structure

BLOCK NAME	BLOCK SIZE	GEOG. BREAKD	AUCTION START	AUCTION END
A	30 MHz	MTA	12/5/1994	3/13/1995
B	30 MHz	MTA	12/5/1994	3/13/1995
C	30 MHz	MTA	12/18/1995	5/6/1996
D	10 MHz	BTA	8/26/1996	1/14/1997
E	10 MHz	BTA	8/26/1996	1/14/1997
F	10 MHz	BTA	8/26/1996	1/14/1997

A, B, and D block auctions were one hundred percent successful. No licenses were in default and reclaimed by the FCC. E and F bandwidth auctions were also highly successful, with few defaults.[10] However, in the C block, with the twin characteristics of 30 Mhz of bandwidth and entrepreneurial status, the default rate was high; eighty percent of this bandwidth was reauctioned.

Speculative bidding was an issue particularly for C block auctions. While both the C and F blocks were entrepreneur blocks, C block spectrum generally commanded a higher price because of its greater bandwidth. Numerous winning bidders defaulted when they lacked the financial backing to build out systems and were unable to resell their licenses profitably. Some winning bids may have been cast on a speculative basis by bidders with no serious intent of building out PCS systems. Several smaller regional bidders did successfully acquire spectrum, mostly 10 Mhz spectrum in non-entrepreneurial blocks or in the 10Mhz entrepreneurial F block, and built out regional wireless service.

Two cases of C block defaults, GWI PCS and NextWave, were litigated up to the U.S. Supreme Court. GWI PCS was the high bidder for fourteen C-block licenses, totaling approximately $1.06 billion. A U.S. bankruptcy court subsequently voided $894 million of GWI's obligations to the FCC, sustaining the argument that, at the time GW I actually took possession of the licenses, the licenses had significantly depreciated in value. The Supreme Court declined to reconsider this case. GWI and its affiliates today are providers of landline phone and internet service.

NextWave was the successful bidder on 86 licenses, mostly C block licenses, for a total of $4.7 billion. When NextWave defaulted, the FCC repossessed the licenses and reauctioned them. Verizon, the top bidder in this auction, agreed to pay $8.8 billion for 133 (now subdivided) licenses. Cingular (now AT&T) and other carriers also successfully bid on this spectrum. In October 2002, the U.S. Supreme Court held that bankruptcy law precluded the FCC from revoking the licenses.[11] In 2004, the FCC and NextWave reached a settlement distributing the spectrum among Cingular, (17 %), NextWave (10%) and the FCC (72%). NextWave's subsequent behavior supports the assertion that its activity in the PCS auctions was primarily speculative.[12]

In 1999, the FCC partitioned some of the C bandwidth into 15 Mhz blocks; for the 2001 auctions of remaining C block spectrum, the bandwidth was further partitioned into 10 Mhz. FCC subdivision of the initial 30 Mhz C block spectrum into narrower bandwidths reflects the nonviability of reserving wide bandwidth (30Mhz) solely for entrepreneurial bidders.

CONSTRAINED MARKET FORCES DETERMINE INDUSTRY STRUCTURE

While the FCC auction structure set an upper limit on the number of PCS competitors, the initial number of potential competitors was determined by the auction outcome. Despite an upper bound of six competitors built into the auction design, the auction outcome, market-driven, was decidedly more concentrated. The two winners of the greatest number of auctions, Sprint/Wireless LP and ATT, achieved penetration rates of 93 and 94 percent respectively. The penetration rate measures the percentage of total MTA's in which a company acquired spectrum. The mean penetration rate of the twenty most successful bidders was 27.4 percent.

In major blocks A and B, the two greatest winners, Sprint and ATT, claimed over fifty percent of total spectrum auctioned; the four greatest winners acquired over seventy percent. In D block auctions, the two largest winners, once again Sprint and ATT, claimed thirty and twenty-five percent of spectrum, respectively. In the E block auctions, the two largest winners, again Sprint and ATT, acquired fifteen and twenty percent, respectively.

Table 4. Initial PCS auction outcomes

Bidder	Spectrum (MHz)	% total spectrum	# MTA's	% total MTA's
WirelesLP	9510	0.163	29	0.6
ATT	8220	0.141	45	0.94
Primeco	3330	0.057	11	0.23
WstnPCS	2310	0.04	21	0.44
Nxtwave	1990	0.034	27	0.56
AmerPort	1890	0.032	6	0.13
Sprint	1570	0.027	16	0.33
GTE	1290	0.022	4	0.08
Powertel	1090	0.019	6	0.13
OPCSEGI	1050	0.018	23	0.48
SWBell	1030	0.018	10	0.21
BellSouth	1010	0.017	8	0.17
AllTel	740	0.013	12	0.25
Ameritech	630	0.011	2	0.04
PacTeles	600	0.01	2	0.04
Omnipt	600	0.01	1	0.02
USWest	420	0.007	9	0.19
Mercury	240	0.004	5	0.1

Bidding across bandwidths was the strategy used by the two major auction winners to assemble a map of broad PCS coverage. Sprint and its partners acquired A or B licenses in 29 MTA's and then acquired 128 D licenses and 31 E licenses in 15 additional MTA's, yielding the Sprint partnership 188 licenses in 45 MTA's. ATT won A/B licenses in 21 MTA's and then acquired 128 D and 107 E block licenses, a total of 256 licenses spanning 45 MTA's. Omnipoint and OPCSE Galloway (a.k.a. Omnipoint) jointly obtained spectrum in approximately half of the MTA's. Two other major bidders each acquired spectrum in about one-half of the MTA's. The two greatest winners, ATT and Wireless LP/Sprint, acquired over 40 percent of the spectrum successfully auctioned. Table 4 displays the PCS auction results for bidders who obtained more than one percent of total spectrum auctioned, or who were successful in at least five auctions.

The auction results were significantly more concentrated than the auction design. The auction design increased the transaction costs of spectrum ownership and the outcome suggests that a more concentrated auction design would have distributed PCS spectrum more efficiently.

EARLY PCS BUILDOUT

The FCC required PCS licensees to make their services available to one-third of the population in their service areas within five years and to two-thirds within ten years.

Major auction winners increased their market shares by purchasing leases in secondary markets and by mergers/acquisitions. In 1998, for instance, three of 1997's top 25 operators in subscribership had consolidated with other carriers.[13] A primary motivation was the desire of larger regional carriers to capture economies of scale and minimize roaming charges, so that they could compete effectively with national operators like ATT and Sprint.

By 1998, fourteen companies had built out PCS service in a total of thirty-seven MTA's [14]. Sprint offered PCS service in 36 MTA's, a penetration rate of 87.8 percent. PrimeCo (now Verizon) provided PCS service in 10 MTA's, and ATT offered PCS service in 15 MTA's. Most built out licenses were in the A and B blocks.

Population and population growth were significant factors determining early PCS build-out.[15] Among BTA's in the top quartile (the most populated BTA's), 98 percent had one or more new entrant providing PCS service, while in the bottom quartile, only 41 percent of the BTA's had PCS coverage.[16] Possible reasons: higher expected profitability from a larger pool of potential customers, and the desire to offset the higher spectrum price in more populated areas[17] with an earlier stream of returns.

ROLE OF MUNICIPAL GOVERNMENTS IN THE DEVELOPMENT OF THE WIRELESS INDUSTRY

Wireless development created a conflict between municipalities' longstanding power to oversee land use and the federal mandate to build out wireless systems. The federal Standard Zoning Enabling Act of 1926 had led to the development of comprehensive zoning ordinances in many municipalities. Zoning ordinances typically define and limit the types of activities and structures allowable in designated districts. These regulations are meant to mitigate the negative externalities that a real estate owner's use of his/her property might impose on other members of the community, and thus to protect property values.

The build-out of PCS infrastructure required the erection of monopoles and towers on which to site wireless antennae, especially in areas lacking tall buildings. The need for towers increased dramatically, both because of the increasing demand for wireless service and because the use of PCS spectrum (located at a higher frequency range) required towers closer together than the older cellular technology.

Communities, especially in urban and suburban areas, opposed tower construction, citing aesthetic and property value concerns. Because the wireless industry relies on radio frequency bands for signal transmission, some municipalities even raised issues of health. The Telecommunications Act of 1996 (TCA) prevented municipalities from denying permission for cell tower construction solely based on health arguments, giving the FCC express regulatory power over wireless facilities' radio frequency emissions. Some municipalities passed ordinances banning cell towers from certain locations, and municipalities began to amend their zoning ordinances to require permitting approval prior to cell tower construction[18]. As of September 25, 1997, tower siting moratoria had been enacted in 202 municipalities in 23 states. These moratoria, enacted primarily in urban areas, affected a population of over six million people.[19] The cell tower controversy has not abated. In 2007, the town of Acton, Ma. voted to temporarily ban new cell tower construction, while the town studied tower placement issues.[20]

The FCC attempted to mediate the conflicting interests of local governments/citizens and the wireless industry. In August 1998, the FCC announced a facilities siting agreement between the Commission's Local and State Government Advisory Committee, the Cellular Telecommunications Industry Association, the Personal Communications Industry Association and the American Mobile Telecommunications Association. To quote FCC Chairman Kennard, "(t)he towers and antennae …are essential …to these new (wireless) technologies. On the other hand, local governments and citizens clearly have a legitimate interest in where and how towers are sited." The joint agreement involved appropriate guidelines for tower and antennae siting, establishing a set of best practices, as well as an informal dispute resolution process for siting issues.[21]

When municipalities have failed to grant permission for tower construction, citing moratoria or zoning ordinances, wireless companies have frequently filed lawsuits to permit construction. Courts have often shown a willingness to favor the federal Telecommunications Act over the wishes of local regulators. In Sprint Spectrum L.P. v. the City of Medina (1996), a federal district court upheld Medina, WA's six-month tower moratorium as an appropriate use of the city's authority to determine telecommunications policy and procedures. However, in Illinois RSA No.3 v. County of Peoria (1997), a federal district court overturned a county's decision denying a plaintiff's request for a cellular tower, ruling that adequate reasons for the denial had not been given and that the mere existence of opposition is insufficient to support denial.[22]

A landmark case regarding cell tower erection is Omnipoint Communications MB Operations LLC v. Town of Lincoln, decided in August 2000.[23] This decision limited a municipality's authority to use a zoning ordinance to limit cell tower construction The U.S. District Court held that the town's "application of (its) zoning ordinance had created (a) substantial gap in coverage within (the) town, and thus had (the) effect of prohibiting provision of wireless services in violation of (the)TCA." The Court found that " A universal technological advance cannot be permanently stayed by a local stratagem devised to preserve the character and beauty of a locality. The enduring principle of local control of land use shall be honored, but only if a federal law which promotes the establishment of a comprehensive nationwide wireless communication network is not thereby subverted." [24] Citing earlier cases, the Court stated that the 1996 Telecommunications Act was passed "in order to provide a procompetitive…framework…to accelerate rapidly private sector deployment of advanced telecommunications and information technologies…. The TCA significantly limits the ability of state and local authority to apply zoning regulations to wireless telecommunications. Under the TCA, the power of local governmental authorities to regulate the placement, construction, and modification of WCF's (wireless communications facilities) is tempered by the proviso that such regulation shall neither discriminate among providers of personal wireless service nor prohibit … the provision of personal wireless service." [25]

WIRELESS INDUSTRY: 2002

By 2002, digital technology was prevalent on the cellular spectrum. The distinction between PCS and cellular service was no longer discernible to the typical consumer.

There had been considerable PCS buildout. Three wireless companies offered PCS service in most or all MTA's. Two of these, Sprint and ATT, had acquired spectrum at auction. The third, VoiceStream, emerged as an industry player through mergers and acquisitions.

Significant vertical integration occurred between 1998 and 2002. Companies merged and otherwise acquired spectrum to build out more nationwide service. Three of the five nationwide wireless companies that were dominant market players in 2002 (VoiceStream, Cingular and Verizon) were the result of mergers and acquisitions since 1998.[26] The permissive attitude of the FCC allowed mergers and acquisitions which dramatically altered the industry's structure from the auction design of the previous decade.

Western PCS (owned by Western Wireless) had successfully bid on PCS spectrum and by 1998, its division, Voicestream, had built out service in several license areas. In 1999, VoiceStream was spun off from Western Wireless. VoiceStream merged with Omnipoint and Aerial (a/k/a American Portable) in 2000 and with Powertel in 2001. In May 2001, Deutsche Telekom AG completed its acquisition of VoiceStream. By 2002, VoiceSteam offered PCS service in 44 MTA's. In September 2002, the VoiceStream brand name was changed to T- Mobile.

Cingular's[27] history originated with the post-Divestiture Baby Bells. Cingular was formed in

2000 as a joint venture of SBC and Bell South, two Baby Bells that had successfully bid on PCS spectrum (Table 4). During the PCS auctions, SBC was called Southwestern Bell. Southwestern Bell changed its name to SBC Communications in 1995. In 1998, SBC acquired Pacific Telesis (a holding company for Baby Bells Pacific Bell and Nevada Bell), which had also acquired PCS spectrum at auction. In 1999, SBC merged with Ameritech, a regional holding company for six Baby Bells. Ameritech was also a PCS auction winner. Despite the increased concentration resulting from this merger, the FCC determined that, with conditions, the SBC/Ameritech merger was in the public interest and would strengthen rather than reduce competition.[28] Consequently, SBC controlled three of the seven regional Baby Bells established under the terms of the Divestiture Agreement with AT&T. The formation of Cingular in 2000 further increased SBC's market position.

Prior to these mergers and acquisitions, the participating Baby Bells had begun to build out PCS licenses in their service areas, and this build out continued under SBC. By 2002, Cingular offered PCS service in thirteen MTA's. However, Cingular's total wireless coverage was much greater. Cingular's Baby Bells, as local landline companies, had acquired cellular licenses and built out cellular networks in the 1980's. According to the FCC, by 2001, Cingular was the second largest provider of wireless services. [29]

Mergers and acquisitions also created Verizon Wireless. Two regional Baby Bells, Bell Atlantic and NYNEX, had merged in 1996 as Bell Atlantic. In 1998, Bell Atlantic and GTE entered into a merger agreement. Bell Atlantic was a provider of cellular service, while GTE had acquired PCS spectrum at auction (Table 4) and had begun to build out service. While the Department of Justice (DOJ) was reviewing the Bell Atlantic/ GTE merger, GTE expanded its wireless licenses by acquiring half of Ameritech's assets.[30]

DOJ opposition to the Bell Atlantic/GTE merger centered around Bell Atlantic's interest in PrimeCo.[31] The DOJ alleged that the merger would reduce competition in nine states. The consent decree required that overlapping wireless service be sold off to other providers.[32] In June 2000, Bell Atlantic acquired GTE, and renamed the merged entity Verizon Communications.

Meanwhile, in 1999, Bell Atlantic and Vodafone Airtouch announced their intention to form Verizon Wireless. Airtouch (Pac Tel Cellular) had just completed a merger with British Vodafone. The DOJ filed a civil antitrust complaint, claiming that the combined effects of the Bell Atlantic/GTE merger and the Bell Atlantic/Vodafone partnership would reduce wireless competition. A revised consent decree required further divestiture of assets in regions where both Bell Atlantic and Vodafone offered wireless service.[33] While the DOJ's consent decrees had the apparent effect of preserving the level of competition in the wireless industry, these mergers in fact created a large wireless provider with considerable market power.

Bell Atlantic's interest in PrimeCo, noted previously, eventually led to the inclusion of PrimeCo in the Verizon group. PrimeCo, formed in 1994 as a joint venture of Bell Atlantic, NYNEX, U.S. West and Airtouch, bid in the PCS auctions, acquiring licenses in eleven MTA's (Table 4). The Bell Atlantic/NYNEX merger, and AirTouch's purchase of the U.S. West shares, resulted in joint ownership of PrimeCo by Bell Atlantic and AirTouch. When AirTouch was purchased by Vodafone, Bell Atlantic exercised its option to dissolve PrimeCo[34] and divide its licenses between itself and Vodafone. As noted earlier, Vodafone and Bell Atlantic then partnered to form Verizon Wireless.

The FCC justified the numerous mergers and acquisitions which it approved during this period, most notably VoiceStream, Cingular and Verizon, on the basis of the economies of scale[35]. These mergers and acquisitions across MTA's resulted

Table 5. PCS and total wireless service provider

Company	#MTA:PCS	PCS PENETRATION RATE	# MTA:total wireless (PCS+cellular)		TOTAL PEN RATE
Sprint	48	1.000	48		1.000
ATT	46	0.958	48		1.000
Voicestr	44	0.917	44		0.917
Cingular	13	0.271	38		0.791
Alltel	1	0.021	3		0.063
Cric/L	12	0.250	12		0.250
Veriz	14	0.292	44		0.917
Qwest	3	0.063	3		0.063
Nextel	0		48		1.000
US Cellular	0		4		0.083
other*(total)	8	0.021	9		0.021
mean		0.246			0.330

in increased vertical concentration of the wireless industry. To quote the FCC in its Fifth Annual CMRS Report, "operators with larger footprints can achieve economies of scale and increased efficiencies compared to operators with smaller footprints. Such benefits permit companies to introduce and expand innovative pricing plans, reducing prices to consumers," intensifying competition among nationwide providers [36].

In November 2001 the FCC raised the spectrum cap for individual wireless companies from 45 to 55 MHz per urban market, and voted to eliminate the cap entirely in January 2003. Raising and eventually eliminating the ceiling on the amount of spectrum for which a provider could acquire leases enabled greater concentration in the wireless industry.[37]

By 2002, the major wireless operators, in descending order, were Verizon, Cingular, AT&T, and Sprint.[38] VoiceStream also had a major nationwide presence, though its number of subscribers trailed the top four. Sprint, AT&T and VoiceStream had built out PCS service in all or most of the MTA's in continental United States. Verizon and Cingular, in contrast, were still providing service to the vast majority of MTA's they serviced using cellular bandwidth, gradually converting this service from analog to digital systems. Nextel also had a nationwide presence in the wireless industry through its ESMR service. Table 5 shows provider penetration rates, by MTA, for both PCS service and total wireless service.

THE WIRELESS INDUSTRY: 2004

By 2004, the marketplace distinction between cellular and PCS service was nonexistent. Service was increasingly digital rather than analog, regardless of the spectrum used for transmission. We continue to discuss PCS buildout because PCS service was the growth segment of the wireless industry.

From 2002 to 2004, buildout of PCS licenses continued[39]. The average number of PCS competitors in each of the 48 primary MTA's increased from 4 in 2002 to 4.6 in 2004. Most of this expansion, however, occurred as cellular providers acquired PCS bandwidth in order to increase their call capacity and provide customers with better coverage and fewer dropped calls. Average total providers per MTA was virtually unchanged: 6.38 in 2002 v. 6.65 in 2004.

While in 2002, eight wireless providers serviced only a single MTA, by 2004, there was no single-MTA provider. Single-MTA providers had either broadened their field of service, discontinued service, or merged/were acquired.

Every PCS provider with a 2002 penetration rate of less than one hundred percent increased its penetration rates from 2002 to 2004. The mean PCS penetration rate rose from 24.6 percent in 2002 to 41.5 percent in 2004. Two of the three dominant PCS providers had 100 percent market penetration rates in 2004, compared with one provider in 2002. Penetration rates also increased in the total wireless industry, driven by increased penetration in the PCS segment. The 2004 mean penetration rate was 60.25 percent, compared to a 2002 penetration rate of 33.02 percent. Four wireless providers had penetration rates of 100 percent by 2004. Several regional companies also provided wireless service.

The acquisitions and mergers that took effect prior to the fall of 2004 were vertical, or interregional, in nature. These acquisitions and mergers had the effect of yielding more companies with nationwide coverage, but did not reduce intraregional competition. As noted previously, the mean number of PCS providers per MTA rose from 1998 to 2004, and the mean number of wireless providers rose through 2002 and then remained relatively constant.

HORIZONTAL MERGERS REDUCE COMPETITION

- **Cingular/AT&T** The nature of wireless mergers changed dramatically in 2004. In February 2004, Cingular, the second largest provider, had entered into an agreement to acquire AT&T Wireless, the third largest wireless provider, spun off from AT&T Corp in 2001.[40] In October 2004, the Justice Department's anti-trust division and the FCC approved Cingular's takeover of AT&T Wireless, to form the largest wireless company in the United States. [41] The settlement required Cingular to divest assets in thirteen markets. According to the Justice Department complaint, Cingular Wireless and AT&T Wireless were two of six nationwide mobile wireless providers; the proposed transaction would reduce competition for wireless telecommunications services in ten markets, and reduce competition for wireless broadband in three additional markets.[42] In nine of these ten wireless telecommunications markets, Cingular and AT&T Wireless were, or held interests in, the two largest incumbent wireless providers, and in all ten markets, the merged firm would be the largest.[43]

The FCC had found that the merger of Cingular and AT&T Wireless resulted in a "sharp drop in the percentage of the population living in counties with more than six providers, (w)hile the percentage of the U.S. population living in counties with three or more, four or more, or five or more mobile telephone carriers (was) unchanged."[44]

Evidence of the decrease in competition resulting from the Cingular/ AT&T Wireless merger is found in a comparison of the HHI's[45] reported in the FCC's Annual CMRS Competition Reports for 2004 and 2005. The FCC computes HHI's for Economic Areas to provide information on the competitiveness of wireless markets in these Areas. Because wireless spectrum was allocated by MSA's and RSA's (for cellular spectrum) and by MTA's and BTA's (for PCS spectrum), the geographic entities for which the DOJ required Cingular/AT&T spectrum divestiture do not align with the geographic entities for which HHI values are available. However, for geographic entities where comparison was possible, we compared pre and post-merger HHI values, as reported by the FCC. In three of the four comparisons, by the FCC's own metric, there was a significant decrease in competition after the merger: Shreveport, LA:

2524 v. 3387; Oklahoma City, OK: 1836 v. 3714; Dallas/Fort Worth Tx: 1743 v. 2708; Topeka KN: 1783 v. 1760.

The FCC did not provide a compelling argument that its approval of this merger was consistent with its Congressional mandate, in the Omnibus Budget Reconciliation Act of 1993, to promote competition in the wireless industry. In the 2005 CMRS Report, the FCC stated: "in industries where the scale of output at which a firm can fully exploit scale economies is large relative to potential demand, there will be room in the market for only a small number of firms operating at the lowest possible costs."[46] The question of whether there are economies of scale of sufficient magnitude to justify this merger was never addressed. The FCC had cited, in its 2004 CMRS Report, the Cingular claim that the combined entity would generate more that $1 billion in operating and capital expenditure savings in 2006 and in excess of $2 billion in annual savings beginning in 2007.[47] The validity of these alleged savings, or their effect on Cingular's pricing, were not reported. The argument for horizontal mergers on the basis of economies of scale is premised upon the realization of cost savings that are passed forward to consumers as lower prices. In his Concurring Statement in the 2005 CMRS Report, Commissioner Copps commented that while Congress tasked the FCC with doing "an analysis of whether or not there is effective competition in commercial mobile services, the failure to define "effective competition" "limits the ability of the Commission and the Congress to rely on our results."[48]

- **Sprint/Nextel** The movement toward horizontal concentration continued. In December 2004, Sprint announced plans to acquire Nextel. On August 12, 2005, having received regulatory approval from the FCC and the DOJ, the companies completed the merger.[49]Sprint Nextel asserted at the time that they were positioning the company for "a seamless integration" of wireless service.

These two mergers/acquisitions had the effect of reducing the average number of wireless providers per MTA by approximately thirty-three percent. The achievement of economies of scale was relatively straightforward for Cingular, since both AT&T and Bell South had selected TDMA technology[50] for their second-generation wireless networks.[51] Sprint, in contrast, had selected CDMA technology[52] for its wireless network,[53] while Nextel used iDEN[54], with included a digital two-way radio feature like a walkie-talkie. Achieving economies has proved more difficult for Sprint/Nextel. While the merger provided the company with scale, Sprint continues to struggle to knit together two very different phone networks (CDMA v. iDEN) and customer groups.[55]

- **AT&T/Bell South** Another horizontal merger that reduced competition was the merger of AT&T with Bell South in 2006. In early 2005, SBC announced its intent to purchase AT&T Corp., a merger that was approved by the DOJ and the FCC and finalized in November 2005. SBC then changed its corporate name to AT&T. In March 2006, AT&T announced its proposed merger with Bell South. Although AT&T and Bell South did not compete directly in the local, long-distance or broadband markets, they competed for business customers. Further, the merger would give AT&T complete ownership of Cingular (a joint venture of SBC and Bell South). On October 11, 2006, the DOJ's antitrust division issued a statement approving the merger unconditionally: "After thoroughly investigating AT&T's proposed acquisition of Bell South, the antitrust division determined that the proposed transaction is not likely to reduce competition substantially. The merger would likely

> result in cost savings and other efficiencies that should benefit consumers."[56]

There was significant opposition to the DOJ ruling. Smaller phone-company rivals, Democrats on the FCC, and consumer groups complained[57] that the DOJ had abdicated its responsibility to scrutinize whether the new company would be so big that it could reduce competition.[58] FCC Commissioner Copps argued that the DOJ had walked out on consumers and small businesses by refusing to impose even a single condition in the largest telecom merger the nation had ever seen.[59] Commissioner Adelstein called the decision "a reckless abandonment of DOJ's responsibility to protect competition and consumers."[60] Most of these concerns focused on wireline phone and broadband services (TV and internet), rather than the wireless segment.

After having delayed its vote,[61] on December 29, 2006, the FCC approved the merger, concluding that significant public interest benefits were likely to result. For the wireless industry specifically, the FCC concluded that benefits would include "improved wireless products, services and reliability due to the efficiencies gained by unified management of Cingular…."[62] As with the previous Cingular merger, the FCC did not explore the magnitude of economies of scale in the wireless industry nor did the FCC define what it considered to be effective competition.

To "facilitate the speediest possible approval of the merger", AT&T had agreed to several voluntary, enforceable commitments.[63] These commitments dealt primarily with broadband, and included a commitment to network neutrality. The primary wireless commitment was the reassignment/transfer to an unaffiliated third party all 2.5 GHz spectrum (broadband radio/educational services) and the commitment to offer service in the 2.3 GHz band to 25% of the population in the service area of AT&T/BellSouth wireless communications licenses.[64]

AWS SPECTRUM: 2006

In 2006, the FCC auctioned 1,122 licenses for 90 megahertz of AWS (Advanced Wireless Services) spectrum at 1710-1750 and 2110-2155 MHz. This spectrum, previously used by federal operators, can be used for a variety of wireless services, including Third Generation (3G) mobile broadband and advanced wireless services.

AWS auctions followed the PCS auction structure of dividing spectrum into six frequency blocks (A-F). PCS licenses were allocated to 50 MTA's and 493 BTA's. AWS licenses were auctioned by twelve multi-state Regional Economic Areas (REAG), 176 Basic Economic Areas (BEA), and 734 highly- disaggregated Cellular Market Areas (CMA). Table 6 shows the structure of the AWS auctions.

Table 6. AWS auction structure

BLOCK NAME	BLOCK G SIZE	EOG BREAKD	# LICENSES	AUCTION START	AUCTION END
A	20MHz	CMA	734	08/09/06	09/18/06
B	20MHz	EA	176	08/09/06	09/18/06
C	10MHz	EA	176	08/09/06	09/18/06
D	10MHz	REAG	12	08/09/06	09/18/06
E	10MHz	REAG	12	08/09/06	09/18/06
F	20MHz	REAG	12	08/09/06	09/18/06

Table 7 displays the ten highest winning bidders, by provisionally winning bets (PWB). [65]

Top Ten Bidders by Provisionally Winning Bids (PWB) Total					
		# licenses			
Bidder	*PWB's*	REAG	BEA	CMA	*PWB Total*
T-Mobile	120	10	17	93	$4,182,312,000
Cellco d/b/a Verizon	13	4	2	7	$2,808,599,000
Spectrum Co LLC	137	1	136	0	$2,377,609,000
Metro PCS	8	2	6	0	$1,391,410,000
Cingular	48	3	24	21	$1,334,610,000
Cricket	99	1	25	73	$710,214,000
Denali Spectrum	1	1	0	0	$365,445,000
Barat Wireless	17	1	5	11	$169,520,000
AWS Wireless	154	1	48	105	$115,503,000
Atlantic Wireless	15	0	14	1	$100,392,000

The AWS auctions offered discounts or bidding credits to smaller businesses that qualified as designated entities. A small business, defined as a bidder with average gross revenues between $15 million and $40 million for the preceding three years, received a 15 percent discount on its winning bid. A very small business, defined as a bidder with average gross revenues not exceeding $15 million for the preceding three years, received a 25 percent discount on its winning bid. Unlike the PCS auctions, however, there were no entrepreneurial blocks limited exclusively to smaller businesses. Moreover, the experience of the PCS auctions showed that larger companies had won entrepreneurial discounts by partnering with smaller firms eligible for discounts. To close this loophole, the FCC approved new rules in April 2006 which extended from five to ten years the period before which a winning smaller business could sell its licenses to another carrier, and placed further restrictions on leasing spectrum or partnering with other carriers. One hundred two designated entities qualified to bid, comprising 61 percent of total bidders. 58 of these qualified as very small businesses and 44 qualified as small businesses.

The AWS auctions were largely successful. Of 1122 licenses auctioned, 1087 licenses were successfully placed. In the REAG auctions, all licenses available in continental United States, Hawaii and Puerto Rico were successfully placed. In the BEA auctions, three licenses were held by the FCC. The vast majority of auction failures occurred in the highly disaggregated CMA auctions.

Of the six top winning bidders, three (T Mobile, Verizon, and Cingular) were wireless companies already offering nationwide service. Spectrum Co. LLC was formed as a consortium of cable owners Comcast, Time Warner, Cox, and Brighthouse, and Sprint. Metro PCS, and Cricket (a/k/a Leap), were established regional providers.

The highest bidder, T Mobile, the fourth largest wireless company, was quoted as saying that its new licenses doubled its average capacity in the top 100 U.S. markets and that it expects to spend nearly $2.7 billion on a network upgrade that can deliver mobile multimedia capabilities that its rivals currently offer.[66] Cellco d/b/a Verizon, the nation's number two wireless provider, acquired sufficient spectrum, according to one analyst, to build a new nationwide wireless network, perhaps devoted exclusively to mobile TV or high speed wireless internet access.[67]

The four last bidders in the top ten list, shown by bidding name, are non-providers and thus seem to have the potential to expand competition. In fact, however, three of these bidders, all of whom qualified for bidding credits as Designated Entities, were affiliated with and/or funded by existing wireless providers.

Denali Spectrum received a 25 percent bidding credit. Denali was funded by Cricket/Leap[68], an existing regional provider that was not eligible for bidding credits. Similarly, Barat was awarded a 25 percent bidding credit. US Cellular, which provides wireless services in Chicago, Milwaukee, San Antonio,and Portland, Oregon, is a limited partner in Barat and provided initial funding for Barat's participation in the AWS auctions.[69] While US Cellular did not acquire any AWS licenses in its own name, the Barat licenses both expanded US Cellular's market area and deepened its infrastructure in its core areas.

AWS Wireless, the top bidder in terms of licenses won, is a wholly owned subsidiary of NextWave Wireless Inc. AWS Wireless did not receive any bidding credits. NextWave Wireless does not currently offer wireless services, nor does its parent corporation, NextWave Telecom Inc. [70]

Atlantic Wireless, which received a 25 percent bidders' credit, is an authorized Verizon dealer in Maryland, New Jersey and Virginia. Moreover, Atlantic is backed by some of the same investors supporting Aloha Partners, a start-up that has major cellular spectrum holdings.[71]

The AWS auction structure, with three overlapping geographic groupings, enabled major providers to acquire additional spectrum without directly competing. The AWS auction results (Table 7) show that this occurred, either coincidentally or collusively. In the REAG auctions, the three largest winners were T Mobile, Verizon, and Cingular. Spectrum Co., on the other hand, focused on the BEA auctions, where it acquired 136 PWB's.

There are two ways in which the spectrum licensed at the AWS auctions potentially could have increased competition in the wireless industry.[72] First, a potential competitor in a region was created when a non-provider in a geographic area won spectrum. Second, an existing provider acquired spectrum, enabling deepening of coverage and upgrades that enhanced the provider's long-term survival.

Although numerous companies were successful bidders, the major AWS winners, discussed previously, were mostly existing wireless providers, or small businesses funded by existing providers. Thus the primary effect of the AWS auctions was a deepening of spectrum for existing providers. However, two regional providers, Metro PCS and Cricket, won spectrum that will enable them to enter major markets such as New York City, Philadelphia, and Washington, D.C.

To examine the effect of AWS spectrum on the number of competitors in each MTA, we mapped the AWS auction results for REAGs and BEAs onto the forty-eight major MTA's[73], the primary geographic designation in the PCS auctions. MTA's are numbered in descending order by population, so wireless data displayed by MTA rank orders the results by population. On average, seventy-eight percent of all AWS spectrum successfully auctioned was acquired by existing providers to deepen their service capacity. On average, seventy- four percent of AWS spectrum was acquired by the three nationwide providers and Spectrum. In every MTA, a majority of AWS spectrum deepened the spectrum held by existing providers, rather than introduced new providers into the market. Further, deepening occurred primarily among major national providers, even when regional competitors were in that marketplace.

Deepening of spectrum holdings by existing providers is not necessarily an undesirable outcome, as discussed earlier. More troubling is the finding that much of the spectrum not acquired by existing providers was won by entities not currently providing service and whose spectrum winnings seem insufficient for them to enter the market as competitors.

Table 8 displays information on CMA and BEA licenses acquired by minor bidders who do not currently provide wireless service. The non-providers in Table 8 each acquired fewer than ten CMA or BEA licenses, and no REAG licenses. Reviewing the first data row, eleven bidders not qualifying for bidding credits each acquired one CMA license. Eight bidders qualifying for a 15% credit each acquired one license, and thirteen bidders qualifying for a 25% credit each acquired one license. Summarizing the first section of Table 8, a total of 69 bidders acquired fewer than ten CMA licenses, and did not win any BEA or REAG licenses. Summarizing the second section of Table 8, a total of eleven bidders had a combined total of fewer than ten CMA and BEA licenses, and no REAG licenses. This means that 80 out of the 104 total bidders in AWS auctions acquired such small slivers of spectrum that their industry participation does not seem viable.

It would seem that this spectrum is mostly being held by speculators. As can be seen from Table 8, in many cases, the price paid for the spectrum was reduced by bidding credits. Such an outcome is inefficient. It deprives the public of access to this spectrum and may enable the speculators to reap future windfall profits.

700 MHZ SPECTRUM: 2008

Auction 73, which began January 24, 2008, auctioned off 62 MHz of spectrum in the 700 MHz band, spectrum now in use by television broadcasters.[74] 1099 licenses in total were auctioned off, in five different blocks, by geographic area. The geographic structure of these auctions parallels the structure of the AWS auctions, with licenses auctioned by Economic Areas, Cellular Market Areas, and Regional Economic Area Grouping.

Table 8. AWS results for bidders acquiring < 10

#CMA licenses won	0 bid credits	15% credit	25% credit			
1	11	8	13			
2	9	7	4			
3	3	3	0			
4	1	2	0			
5	0	1	1			
6	1	1	0			
7	2	1	0			
8	0	0	0			
9	1	0	0			
#BEA/CMA licenses won						
1BEA/0CMA	0	0	1			
2BEA/0CMA	0	0	1			
4BEA/0CMA	1	0	0			
1BEA/2CMA	0	1	1			
1BEA/3CMA	1	0	0			
2BEA/3CMA	1	1	0			
2BEA/6CMA	1	0	0			
3BEA/2CMA	1	0	0			
3BEA/5CMA	0	1	0			

The system of small business bidding credits for Auction 73 is the same as the system used in the AWS auctions. The 700 MHz spectrum is particularly desirable because signals broadcast in this spectrum travel relatively long distances and are not so susceptible to interference as higher-frequency waves.

This spectrum block has the potential to increase the level of competition in the wireless industry. The auction plan included a $4.6 billion threshold or reserve price[75], which when met, triggered open-access rules for one-third of the spectrum. The licensee for Band C, the widest block at 22 MHz, "will be required to provide a platform that is more open to devices and applications. This would allow consumers to use the handset of their choice in this spectrum block, subject to certain reasonable network management conditions that allow the licensee to protect the network from harm."[76] The FCC explained why it was imposing open-access rules: "the 700 MHz spectrum provides an important opportunity to apply requirements for open platforms for devices and applications for the benefit of consumers…. (H)owever, … it would not serve the public interest to mandate … requirements for open platforms … for all …700 MHz spectrum or to impose broader requirements, such as wholesale or interconnection requirements, for the C block." Continuing the justification, "while it is easy for consumers to differentiate among providers by price, most consumers are unaware when carriers block or degrade applications, …thus making it difficult for providers to differentiate themselves on this score."[77]

CONCLUDING REMARKS

The wireless telephony market is a dynamic market. Constantly evolving technologies have driven the industry's expansion. For these reasons, it has been challenging for the FCC to implement its Congressional mandate to promote competition in the industry. The FCC has structured spectrum auctions with the goal of promoting competition. However FCC approval of horizontal mergers has had the effect of reducing the number of wireless competitors.

At the municipal level, The Telecommunications Act has prevented cities and towns from limiting the number of wireless providers.

As regional providers who have acquired AWS or 700 MHz spectrum build out infrastructure, in the short run at least, there will be an increase in competition in areas where these regional providers have expanded their footprint. Metro PCS, for instance, is currently in the process of obtaining the necessary municipal approvals to build out AWS spectrum licenses in the Northeast.

Several factors may affect government's future role in the evolving structure of the wireless industry. Auction structures for the PCS, AWS and 700 MHz auctions have allowed and in some instances encouraged speculation on wireless spectrum, creating a private-sector secondary market for the resale of spectrum leases. Such speculation creates market inefficiencies, denying consumers use of this spectrum, probably reducing competition and driving up wireless prices, and providing windfall profits to secondary-market license owners. The market structure of the industry would be improved if the FCC followed the example of many European regulators and subdivided available spectrum into larger units, rather than fragmenting it so greatly.

Another issue is FCC implementation of its mandate to provide a competitive wireless marketplace. In the recent past, the FCC has justified both vertical and horizontal mergers that have reduced competition on the basis of economies of scale. The FCC uses the HHI index to measure competition. It would be useful if the justification of economies of scale were supported by some analysis of the cost structure of the wireless industry. Do economies of scale justify a reduction in the number of providers to three firms, to two firms?

How do Congress and the FCC define "effective competition?" The parameters for determining if a proposed merger reduces competition might be expanded to include the Grinnell Test[78], a test that is typically used by the U.S. courts when adjudicating antitrust cases. Another possibility worth considering, given the dynamic nature of the industry, is examining whether barriers to innovation exist in the wireless market.[79]

Finally, decisions made at the federal, state, and local levels regarding the use of WiFi and WiMax technologies [80]will influence the future structure of the wireless industry. WiFi currently uses license-exempt spectrum. Whether WiMax will ultimately use licensed or unlicensed spectrum is still undetermined. A related issue is whether these new wireless technologies will be provided by existing providers, or whether the regulators will allow a further consolidation of the market, with broadband providers providing a bundle of services which includes wireless telephony.

REFERENCES

American Bar Association (editors). (2005). *Telecom Antitrust Handbook (American Bar Association Section of Antitrust Law Monograph).*

Boston Globe. (2007). *Acton to vote on cell tower moratorium.*

Cellular News. (2006). *DOJ Approves AT&T-Bell South Merger.*

Coase, R. H. (1959). The Federal Communications Commission. *Journal of Law and Economics, 2*, 1-40.

Department of Justice (1999). *Justice Department Requires Bell Atlantic and GTE to Divest Wireless Businesses in order to Proceed with Merger.*

Department of Justice (2004). *Justice Department Requires Divestitures in Cingular Wireless's Acquisition of AT&T Wireless.*

Dunbar, J. (2006). *FCC Delays Vote on AT&T-BellSouth Deal.* Associated Press.

Electrical Engineering Times. (2007). FCC spells out rules for 700 MHz bands. *EE Times.*

FCC News (1998). *Chairman William E. Kennard Announces Historic Agreement by Local and State Governments and Wireless Industries on Facilities Siting Issues.*

FCC News (1999). *FCC Approves SBC-Ameritech Merger Subject to Competition-Enhancing Conditions*

FCC News (2001). *FCC Announces Wireless Spectrum Cap to Sunset Effective January 1, 2003.*

FCC News (2006). FCC Approves Merger of AT&T Inc. and BellSouth Corporation.

Federal Communications Commission. (1998). *Third Annual CMRS Competition Report.*

Federal Communications Commission. (1999). *Fourth Annual CMRS Competition Report.*

Federal Communications Commission. (2000). *Fifth Annual CMRS Competition Report.*

Federal Communications Commission. (2002). *Seventh Annual CMRS Competition Report.*

Federal Communications Commission. (2004). *Ninth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services.*

Federal Communications Commission. (2005). *Tenth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services.*

Federal Communications Commission . (2007) *Second Report and Order, 700 Mhz Bands.*

Financial Times (2008). *Sell phone.*

Harris, N. (2000). The Basic Question Which wireless standard is best? A lot hangs in the balance. *The Wall Street Journal.*

Harvard Law Review (Eds) (2001). Antitrust and the Information Age:Section 2 Monopolization Analyses in the New Economy. *Harvard Law Review, 114,* 5.

Hazlett, T. W. (1998). Assigning Property Rights to Radio Spectrum Users: Why Did FCC License Auctions Take 67 Years? *Journal of Law and Economics*, *41*, 45-67.

Industry Canada. (2007). *Consultation on a Framework to Auction Spectrum in the 2 GHz Range including Advanced Wireless Services.*

McDonough, C. (1999). The Price of Zoning Revisited: Zoning Issues Raised by the Telecommunications Act of 1996. *Illinois Real Estate Letter, 13*(1), 1-3.

McDonough, C. (2005) *History of the Wireless Industry in the United States,* report.

McDonough, C. (2007). *The Impact of AWS on the Structure of the Wireless Telephony Industry*. Paper presented at the 2007 Eastern Economics Association Conference, New York, New York.

McDonough, C. (2008). *The Structure of the United States' Wireless Telephone Industry:Theory, Reality and the Role of Government Regulation*, International Telecommunications Society 17th Biennial Conference, Montreal.

New York Times (1999). *Holy War Over the Future of Wireless.*

Regli, B. J. W. (1997). *Wireless: Strategically Liberalizing the Telecommunications Market.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Sprint Nextel News Release (2005). *Sprint Nextel Completes Merger.*

Sterling, C., Bernt, P., & Weiss, M. (2006). *Shaping American Communications.* New York, N.Y.: Lawrence Erlbaum Associates.

United States District Court for the District of Columbia (2000). *United States of American v. Bell Atlantic, GTE and Vodafone Airtouch, Motion for Entry of Final Judgment.*

Webb, W. (2001). *The Future of Wireless Communications.* Boston, MA: Artech House.

Westlaw. *High Plains Wireless, L.P., Appellant v. Federal Communications Commission, Digital PCS, LLC and Tritel Communications, Inc., Intervenors.*

Westlaw. *Omnipoint Communications MB Operations, LLC v. Town of Lincoln D Mass, 2000, 107 F.Supp.2d*.

KEY TERMS

Antitrust Laws: Laws enacted by the federal government to limit and control concentrations of economic power. The first antitrust law in the United States, the Sherman Antitrust Act, was enacted by Congress in 1890. Antitrust laws are enforced by the U.S. Department of Justice. Antitrust laws prohibit restrictions of free trade and competition, outlaw anti-competitive practices such as price fixing, and limit the scope of mergers and acquisitions.

Competition: A market condition where price and output are determined by the collective market forces of supply and demand, rather than by an individual producer or group of producers. Among the factors that tend to lead to competition are ease of entry into an industry, number of competitors and market shares, and the substitutability among the products of the competing firms.

Market Structure: The level of competition in an industry, that is, whether numerous firms compete with each other for market share or the industry is highly concentrated with few competitors.

Mergers and Acquisitions: The buying, selling and combining of different companies, often companies that produce similar or competing products or services. This is a tool used to enable a company to grow rapidly without having to construct an expanded business entity. Mergers and acquisitions typically reduce the amount of competition in an industry. The difference between a merger and an acquisition is that a merger is often undertaken consensually, while an acquisition may be the result of a takeover of one company by another.

Wireless Spectrum: Bands within the electromagnetic spectrum that have been allocated for wireless communications, such as cellular bandwidth at 800-900 MHz and the PCS bandwidth in the 2 GHz range.

Wireless Telephony: Telephone service which uses radiowaves, rather than a landline, to connect the person initiating a call to a telephone network.

Wireline Telephony: Telephone service which uses copper lines, or fiberoptic cables, to connect the person initiating a call to a telephone network.

ENDNOTES

[1] Regli, B.J.W. (1997). *Wireless: Strategically Liberalizing the Telecommunications Market.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

[2] 2 Annual CMRS Competition Report. Federal Communications Commission, July 3, 2002.

[3] Consultation on a Framework to Auction Spectrum in the 2 GHz Range including Advanced Wireless Services, Industry Canada, February 2007.

[4] Coase, R.H. (1959). "The Federal Communications Commission," Journal of Law and Economics, 2, 1-40.

[5] Hazlett, T.W. (1998). "Assigning Property Rights to Radio Spectrum Users: Why Did FCC License Auctions Take 67 Years?" Journal of Law and Economics, 41, 45-67.

[6] Some of this information on PCS auctions was presented in McDonough, C. (2008), *The Structure of the United States' Wireless Telephone Industry:Theory, Reality and the Role of Government Regulation*, International Telecommunications Society 17th Biennial Conference, Montreal.

[7] The winner's curse refers to the tendency for the winner to be the bidder who most overestimates the value of the item up for bid.

[8] "Reflexive bidding" is a tactic for deterring would-be competitors from bidding on a particular license by encoding bids. See High Plains Wireless, L.P., Appellant v. Federal Communications Commission, Digital PCS, LLC and Tritel Communications, Inc., Intervenors.

[9] FCC 94-178 Federal Communications Commission, Washington D.C. July 15, 1994.

[10] In the E block, there were six defaults (1%) and in the F block, there were two defaults (.4%)

[11] Supreme Court of the United States, Federal Communications Commission v. Nextwave Personal Communications, Inc et. al., October term, 2002.

[12] In April 2005, NextWave reorganized and sold NextWave Telecom, and its PCS spectrum, to Verizon for $3 billion. NextWave Wireless remains a corporate entity, and through its wholly owned subsidiary, AWS Wireless, successfully bid on spectrum in the AWS and other subsequent spectrum auctions. In April 2008, NextWave Wireless announced plans to put up for sale 223 licenses, including its AWS licenses, which together cover about eighty-four percent of the United States. (NextWave Plans to Sell Wireless Spectrum Positions, CNN

Money.com, Dow Jones Newswires, April 23, 2008.)

13 Fourth Annual CMRS Competition Report, Federal Communications Commission, June 24, 1999

14 Buildout information for the 41 largest MTA's was obtained from these sources: wireless dimensions.com and data from providers' websites.

15 McDonough, C. (2005) History of the Wireless Industry in the United States, report.

16 Fourth Annual CMRS Competition Report, op.cit.

17 For example, the high bids for A/B spectrum were $443 million in MTA 1, $198 million in MTA 11, and $34 million in MTA 23. (MTA's are ranked by population in descending order.)

18 Typical of such regulation is the Andover Massachusetts Zoning Bylaw's article (enacted October 2001) requiring the issue of a special permit for wireless communications devices. Included are design standards regarding setback, height, construction, and landscape, and an encouragement of co-location.

19 Third Annual CMRS Competition Report, June 11, 1998, citing CTIA Siting Report Statistics, September 27, 1997.

20 "Acton to vote on cell tower moratorium," Boston Globe, May 6, 2007.

21 "Chairman William E. Kennard Announces Historic Agreement by Local and State Governments and Wireless Industries on Facilities Siting Issues," FCC News, August 5, 1998,

22 McDonough,C. (1999) "The Price of Zoning Revisited: Zoning Issues Raised by the Telecommunications Act of 1996," Illinois Real Estate Letter, 13(1), 1-3.

23 Omnipoint Communications MB Operations, LLC v. Town of Lincoln D Mass, 2000, Westlaw, 107 F.Supp.2d 108.

24 Ibid.

25 Ibid.

26 This history of the formation of VoiceStream, Cingular, and Verizon was presented in McDonough, C. (2008), op.cit.

27 Cingular subsequently became the new AT&T.

28 FCC News, "FCC Approves SBC-Ameritech Merger Subject to Competition-Enhancing Conditions," October 6, 1999.

29 Seventh Annual CMRS Competition Report, Federal Communications Commission, July 3, 2002.

30 SEC, Form 8-K, "GTE to Acquire Ameritech Wireless Assets in Midwest," April 5, 1999.

31 Editors of ABA, "Telecom Antitrust Handbook(American Bar Association Section of Antitrust Law Monograph)", 2005.

32 Department of Justice, "Justice Department Requires Bell Atlantic and GTE to Divest Wireless Businesses in order to Proceed with Merger," May 7, 1999.

33 United States District Court for the District of Columbia, United States of American v. Bell Atlantic, GTE and Vodafone Airtouch, "Motion for Entry of Final Judgment" March 20, 2000.

34 Telephonyonline, August 9, 1999

35 Fourth Annual CMRS Competition Report, Federal Communications Commission, June 24, 1999.

36 Fifth Annual CMRS Competition Report, Federal Communications Commission, August 18, 2000.

37 "FCC Announces Wireless Spectrum Cap to Sunset Effective January 1, 2003," FCC News, November 8, 2001.

38 Seventh Annual CMRS Competition Report, Federal Communications Commission, July 3, 2002.

39 Data on PCS service providers, as of April 2004, were obtained primarily from wirelessadvisor.com

40 Department of Justice, "Justice Department Requires Divestitures in Cingular Wireless's Acquisition of AT&T Wireless," October 25, 2004.

41 Although DOJ anti-trust authorities required Cingular to offload assets in eleven states to ensure adequate competition in these markets, the overall effect of the acquisition was to reduce the level of intraregional competition

42 Department of Justice and Network World (http://www.networkworld.com/edge/news/2004/1025/cingudojr.html)

43 Ibid.

44 Tenth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services, Federal Communications Commission, September 26, 2005.

45 The Herfindahl-Hirschman index (HHI) is calculated by summing the squares of the individual market shares of all firms competing in a market. When there is a single supplier, HHI takes a value of 10,000. The value of the HHI falls (to approach zero) as the structure of the market becomes more competitive.

46 Tenth Annual Report, op.cit.

47 Ninth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services, Federal Communications Commission, September 9, 2004.

48 Tenth Annual Report, op.cit.

49 Sprint Nextel Completes Merger, News Release, Sprint Nextel, Aug. 12, 2005.

50 TDMA (time division multiple access) technology uses relatively narrow batches of wireless frequencies, which it then breaks up many times a second to convey many messages.

51 Harris,N. (2000) "The Basic Question Which wireless standard is best? A lot hangs in the balance," The Wall Street Journal, September 18, 2000.

52 CDMA (code division multiple access) technology breaks messages into small bits, which are then scattered throughout the wireless telephone spectrum and reassembled at their destination.

53 "Holy War Over the Future of Wireless, New York Times, February 15, 1999.

54 iDEN (Integrated Digital Enhanced Network) is a mobile telecommunications technology which provides its users the benefits of a two-way radio and a cellular telephone.

55 "Sell phone," Financial Times, January 21, 2008.

56 M. Reardon, CNET News, October 11, 2006, quoting Thomas Barnett, assistant attorney general in charge of the Justice Department's antitrust division.

57 A.Schwartzman, CEO of Media Access Project, commented " AT&T, with the help of a complicit government, is poised to control half the nation's phone lines and also be the largest wireless … company. (Wait) until (consumers) get their next phone bill." M.Reardon, CNET News, op.cit.

58 "DOJ Approves AT&T-Bell South Merger", Cellular News, 10/12/2006.

59 Ibid.

60 Dunbar, J. (2006) "FCC Delays Vote on AT&T-BellSouth Deal," Associated Press, October 13, 2006.

61 Ibid.

62 FCC News, "FCC Approves Merger of AT&T Inc. and BellSouth Corporation," December 29, 2006.

63 Letter from Robert W. Quinn, Jr., Senior Vice President, Federal Regulatory, at&t, to M.L. Dortch, Secretary, Federal Communications Commission, December 28, 2006.

64 Attachment, "Merger Commitments," Letter from Robert W. Quinn, Jr., op.cit.

65 FCC website, "FCC Advanced Wireless Services Auction No.66 ***Final***

66 Yahoo Finance, AP, Friday October 6, 2006

67 Business Week online, September 7, 2006

68 A Form 8-K was filed by Leap Wireless International, Inc., with the SEC on July 28, 2006, just prior to the start of the AWS auctions on August 9, 2006. Leap Wireless, operating under the Cricket brand, is a wireless provider in several markets including Pittsburgh, Denver, Phoenix, Memphis and Birmingham. Form 8-K reads: "On July 14,2006, Cricket …, a wholly-owner subsidiary of the registrant, Leap…, entered into a Credit Agreement … by … Cricket …as Lender, Denali Spectrum License, LLC …as Borrower, and Denali Spectrum, LLC …as Guarantor. The Borrower is a wholly owned subsidiary of the Guarantor. …Cricket has agreed to make loans to Borrower of up to $203.8 million to fund the payment of the net winning bids for licenses for which the Borrower is the winning bidder. Cricket has also agreed to loan the Borrower an amount equal to $1.50 times the aggregate number of potential customers covered by all licenses for which the Borrower is the winning bidder, to fund the construction and operation by Borrower and its subsidiaries of wireless networks using such licenses. …Cricket has made an equity investment of approximately $7.6 million in the Guarantor, comprising an 82.5 percent non-controlling membership."

69 US Cellular's Form 8-K (September 22, 2006) reports that US Cellular has made capital contributions and advances to Barat Wireless of $79.9 million to provide initial funding of Barat's participation in the AWS auctions and that US Cellular expects to make additional contributions to Barat.

70 NextWave Telecom was founded in May 1995 (as the PCS auctions were being held), as the parent company of NextWave Personal Communications (PCI), NextWave Wireless and TELE*code Inc. NextWave PCI the successful bidder for 63 licenses in the PCS auctions, for a total of $4.7 billion dollars. After paying $500,000, in 1998, NextWave Telecom defaulted and filed for bankruptcy protection. The FCC repossessed the licenses and reauctioned them for nearly $16 billion. The US Court of Appeals for the District of Columbia Circuit ruled that the FCC had no right to cancel NextWave's licenses and the U.S. Supreme Court upheld the lower court's ruling that U.S. bankruptcy law precluded the government from revoking the licenses. In April 2004, the FCC and NextWave Telecom reached a settlement by which the FCC retained funds from the sale of seventeen percent of the spectrum to Cingular, seventy-two percent of the spectrum (1550 MHz of spectrum in 60 markets) was returned to the FCC, and NextWave retained ten percent (300 MHz of spectrum in 25 markets). In April 2005, NextWave Telecom reorganized and sold NextWave Telecom (and the PCS spectrum held by this entity) to Verizon for $3 billion. NextWave Wireless remains a corporate entity and AWS Wireless is a wholly owned subsidiary of NextWave Wireless.

71 http://weblogs.jupiterresearch.com/analysts

72 McDonough, C. (2008) op. cit.

73 McDonough, C. (2007, February). *The Impact of AWS on the Structure of the Wireless Telephony Industry.* Paper presented at the 2007 Eastern Economics Association Conference, New York, New York.

74 This spectrum will become available once these broadcasters switch to digital signals in February 2009.

75 On February 1, 2008, an unidentified company bid $4.7 billion yesterday for the upper 700 MHz C block spectrum, exceeding the

reserve price of $4.6 billion that triggered open-access rules. FCC website, Wireless, Auction 73.

76 "FCC spells out rules for 700 MHz bands," EE Times, 8/1/2007.

77 Second Report and Order, 700 Mhz Bands, Federal Communications Commission, adopted July 31, 2007.

78 The Grinnell Test examines (1) the possession of monopoly power in the relevant market and (2) the willful acquisition or maintenance of that power as distinguished from growth or … historic accident.

79 (2001). Antitrust and the Information Age: Section 2 Monopolization Analyses in the New Economy. *Harvard Law Review,* 114, 5.

80 WiFi technology is defined by IEEE Spec. 802.11. WiMax technology is defined by IEEE Spec. 802.16

Chapter VI
Shaping Regulation in the Australian Mobile Industry

Indrit Troshani
University of Adelaide, Australia

Sally Rao Hill
University of Adelaide, Australia

ABSTRACT

The adoption and diffusion of mobile services has achieved a spectacular growth in many countries around the world. However, regulators in these countries are finding that existing regulatory frameworks are not suitable for dealing with these services. This chapter employs qualitative evidence to investigate how regulation and legislation can affect mobile services in the Australian mobile telecommunications industry and draws from it to propose an innovative regulatory framework. The framework is comprised of five major components: consumer and intellectual property protection, market and resources access, and environmental protection. We argue that these encompass the interests of the stakeholders operating in mobile industry and given its dynamic and complex nature co-regulation is an effective approach that can be used to minimize costs and enhance compliance.

INTRODUCTION

With the advent of broadband wireless networks, the number of 3G subscribers is predicted to rise to nearly 800 million worldwide by the end of 2010 from 70.6 million recorded at the end of 2005 (Informa, 2005b). In addition, many markets have surpassed the 100% handset penetration and in those markets mobile users are attracted to new and compelling mobile services (Mylonopoulos & Sideris, 2006; Steinbock, 2005). These market conditions provide a conducive environment for delivering increasingly sophisticated mobile services which constitute a substantial source of potential revenue growth for the stakeholders of the mobile telecommunications industry (Alahuhta, Ahola, & Hakala, 2005; Kumar, 2004; Massey, Khatri, & Ramesh, 2005).

A mobile service can be defined as an activity or series of intangible activities that occur when mobile consumers interact with systems or service provider employees with the support of a mobile telecommunications network (Bouwman, Haaker, & De Vos, 2007; Van De Kar, 2004). Examples of mobile services include mobile e-mail, SMS and MMS services, content downloads, mobile ticket reservations, mobile stock trading, and mobile TV (Bina & Giaglis, 2005). Mobile services are complex and require the integration of diverse technological and organizational resources which typically cannot be found within a single organization. Consequently, the knowledge necessary for developing and deploying these services may involve several heterogeneous stakeholders who are often embedded in various technological, economic, and social settings. In order to succeed, these stakeholders must interact with each other while complying with legal and societal requirements that balance their diverging interests, motivations, and needs (Camponovo & Pigneur, 2003). These requirements constitute a regulatory regime which, by definition, can operate at either industrial, national or international levels and can influence, direct, limit or prohibit any activity undertaken by stakeholders operating in the mobile industry (Yoo, Lyytinen, & Yang, 2005). Typically, regulatory regimes are set by regulatory and legislative authorities including government agencies, industry and consumer associations.

Combined with the complexity of stakeholder interactions, regulation has the potential to affect the offerings and the uptake of mobile services (Kijl *et al.*, 2005; Sangwan & Pau, 2005). Credible and transparent regulatory rules can boost investments in the industry, promote public confidence and the development of innovative and affordable mobile services while stimulating industry research and development efforts (Verikoukis *et al.*, 2006). However, regulation can also impact the industry in a negative way. Increasing regulatory compliance fees for stakeholders can increase the overall cost of operation well beyond that of acquiring the technology itself (Tongia, 2007).

As the mobile services industry matures, it is attracting the attention of policy makers and consumer advocacy groups that are increasingly focusing on its regulation (Woolfson, 2005). MacInnes (2005) argues that when a service technology "has reached some minimum standard of performance and reliability one needs to be concerned with societal, legal, and general economic factors. This is a stage that is generally overlooked because it is not until a technology has been commercialized that the originators realize the problems that the technology poses to society in general. … A successful implementation of a business model around new technologies would have to consider those potential obstacles and provide solutions." (p. 7).

That is, solutions are needed for potential legal, societal, and general economic concerns that mobile services may introduce (MacInnes, 2005). Regulatory regimes around the globe are ill equipped for dealing with technologies such as mobile services because the "existing policy frameworks have been inherited from specific national, regional, and international histories of regulating broadcasting, telecommunications, and media, as distinct entities, and are not well-placed to deal with contemporary communications technologies that blur the boundaries among these." (Goggin & Spurgeon, 2005, p. 181). Although limited studies about regulatory regimes are available in the literature they lack comprehensiveness (Li & McQueen, 2008; Oh *et al.*, 2008; Tallberg *et al.*, 2007; Buerkler, 2005; Grzybowski, 2005; Han *et al.*, 2006; Jho, 2007). In fact, many scholars argue that research in this area is lacking (Gelenbe, 2003; Killström *et al.*, 2006; Pitkänen, 2006; Ubacht, 2004). Further, practitioners believe that a healthy regulatory environment is essential for the growth of these industries (Chen, 2008; Hsieh, Jones, & Lin, 2008; Informa, 2005b, 2006a; Verikoukis *et al.*, 2006; Woolfson, 2005; Goggin & Spurgeon, 2005). With this study, we attempt to address these concerns and the lack of vigorous research into regulatory frameworks. Our objective is to leverage on extant literature and use qualitative evidence collected to explore how regulation can affect the mobile industry. The study will culminate with a regulatory framework for the mobile services industry. The Australian mobile services industry was used as a setting for investigating this objective.

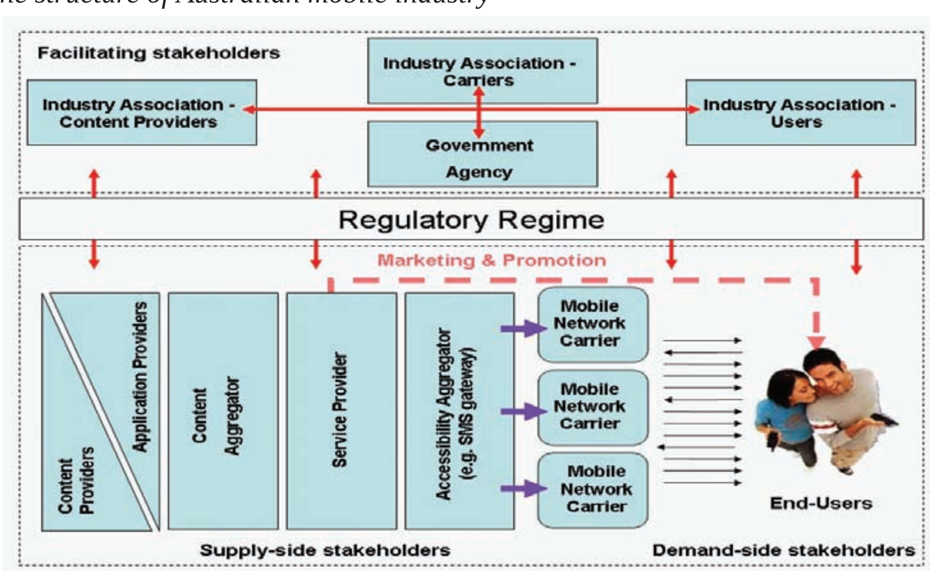
This study is important for several reasons. First, it can be invaluable to stakeholders in the mobile industry in helping them improve their understanding of the environmental factors that enhance or constrain their positions in their value chain, and industry. A deeper understanding of such factors can help stakeholders in many ways in i) achieving a valuable competitive advantage. Stakeholders that exhibit compliance with regulatory rules that benefit users of mobile services may achieve their trust more effectively than those who do not (Killström *et al*., 2006); ii) providing stakeholders the opportunity to "achieve knowledge on legal issues, to stay away from legal areas in which processes are unclear, and to avoid related risks." (Kijl, 2005, pp. 66-67) which decreases potential transaction costs (Kijl *et al*., 2005; Woolfson, 2005); iii) helping avoid unbalanced legal rights amongst stakeholders which can severely threaten businesses by causing otherwise innovative business models to be illicit (Kijl *et al*., 2005; Killström *et al*., 2006; Woolfson, 2005). Second, regulatory and legislative influences can have direct implications on how mobile services and business models are designed and how they operate at organizational, industry and environmental levels. Further, these influences can determine the nature of mobile services that can be offered and their diffusion trajectories amongst end-users (MacInnes, 2005; Spiller & Zelner, 1997).

This chapter is organized as follows. First, we overview the Australian mobile industry, then, we discuss the research method before elaborating the proposed regulatory framework. Finally, we conclude the paper and identify and future research directions.

THE AUSTRALIAN MOBILE INDUSTRY: AN OVERVIEW

To set the scene, we provide a brief overview of the Australian mobile industry. The diagram in Figure 1 has been adapted from the work of

Figure 1. The structure of Australian mobile industry



Troshani and Rao (2007a) and summarises the current structure of the Australian mobile industry. To distribute mobile services to demand-side stakeholders, intense interactions need to occur amongst supply-side stakeholders. These occur in an environment which is being co-regulated by both the industry itself and facilitating stakeholders including the government regulator and various industry associations (Troshani & Rao, 2007a).

The interactions that occur amongst supply-side stakeholders include mobile service development, marketing, distribution and billing. The roles of the major stakeholders shown in Figure 1 are briefly described in Table 1.

A regulatory regime is meant to stimulate the changing market environment from monopoly to competitiveness, protect consumers, but also to be reactive to market dynamics in order not to over- or under-regulate. On the one hand, over-regulation can bring several adverse outcomes to the industry including, high engagement costs and possible duplicative and confusing rules for all stakeholders including end-users (AMTA, 2005a). On the other hand, under-regulation can also lead to adverse outcomes for the industry including unfair market dominance of players which may lead to uncompetitive conditions as well as end-user exposure to unfair and illicit trading practices. The relationship between the regulators, supply- and demand-side stakeholders should be dynamic, mutually responsive, and co-evolutionary, in that the regulatory authorities should match their activities to the issues that

Table 1. Categories of the stakeholders and their roles

Mobile Network Carriers *(also known as mobile network operators) supply a variety of telecommunication services to subscribers who use their network infrastructure. Typically, carriers own network infrastructures or hold licenses for spectrum allocations which are required for providing mobile services to subscribers. Carriers also establish technical support layers and generally have control over subscriber information and billing. This category also includes Mobile Virtual Network Operators (MVNO) which do not own their network infrastructure, but have business arrangements with carriers to buy minutes of use (MOU) to sell to their end-users.*
Accessibility/Connectivity Aggregators *typically operate SMS gateway infrastructures that connect to many carrier networks. They purchase end-to-end mobile services from the carriers and act as an interface between these and service providers.*
Service Providers *provide the technology platform for delivering mobile content to end-users by setting up SMS short codes which can be advertised and promoted through various media properties (e.g. TV and magazines).*
Content aggregator/publishers *bundle content and services from multiple sources. They provide wider dissemination by negotiating intricate, complex and time-consuming agreements with network carriers. Content aggregators also provide mobile content platforms which integrate numerous content sources into a single interface for end-users.*
Content Intellectual Property (IP) owners/labels/content providers/media owners *create original content from brands, traditional music, films, TV, games etc., and deliver it to consumers via network carriers, service providers or content aggregators.*
Application providers *The availability of mobile services is supported by the work of application developers who employ industry development tools to create innovative applications for packaging content offerings.*
Industry Association - Content IP owners/labels/content providers *are responsible for creating a framework for uniting dynamic and diverse industry stakeholders, hosting networking events, business forums, industry development programs, training sessions, advising government on policy development and promoting industry innovation.*
Industry Association - Carriers *represents the mobile carriers that operate in Australia. They ensure that the mobile carrier industry as a whole addresses community and government expectations. They also coordinate the deployment of mobile networks, enhance communication and consultation among stakeholders, and liaise with government.*
Industry Association - Users *represent end users and provide independent views on mobile telecommunications policy and service issues, including competition, regulation, end-user protection.*
Government Agency *is responsible for promoting regulation and competition in the mobile telecommunications industry while protecting consumers. They also aim at creating an environment where community standards are respected and user concerns are addressed.*
Mobile Subscribers *should be considered as a legitimate stakeholder in the mobile telecommunications industry, because their lives are likely to be affected by mobile services.*

arise due to the complex dynamics of the mobile market on an ongoing basis. The manner in which stakeholder interactions are influenced by the regulatory regime has direct implications on the mobile industry, and thus, constitutes the scope of this investigation the method of which is described next.

METHOD

This chapter is exploratory in nature and utilizes qualitative evidence. We are concerned with the extent to which interpretations of regulation are sensible in the Australian context and the way in which these influence the adoption and diffusion of mobile services in Australia (Neuman, 2000). A qualitative exploratory approach enables the study of dynamic, intricate, and multifaceted processes as well as the exploration of emerging themes (Cassell & Symon, 1994; Smith & Fischbacher, 2005; Yin, 1994). Qualitative empirical data were collected via semi-structured face-to-face interviews, which were used because of their flexibility. In-depth interviews provide rich insights for exploring, identifying, and understanding viewpoints, attitudes, and influences (Healy & Perry, 2000). Moreover, they also allow greater control over the interview situation (e.g. sequencing of questions) while providing the opportunity for making clarifications and collecting supplementary information (Frankfort-Nachmias & Nachmias, 1996).

As indicated in the previous section, stakeholders of the mobile service industry can be categorized into eight groups – carriers, accessibility aggregators, content aggregators, service providers, content owners/providers, application providers, industry associations, and regulatory bodies. Industry associations can be broken down

Table 2. Categories of stakeholders interviewed

Interviewee Role	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Carrier						X										X						
Accessibility Aggregator											X							X				
Content Aggregator	X	X																	X	X	X	
Service Provider																						X
Content Owner		X	X		X					X				X		X	X		X	X	X	
Application Developer	X	X	X	X	X		X	X		X				X		X					X	
Industry Association (Content Providers)									X													
Industry Association (Users)												X										
Industry Association (Carriers)															X							
Government Regulator													X									

further into user groups, content providers' association and carriers' association.

The contact details of these organizations were obtained from the Australian Interactive Media Industry Association's (AIMIA) and Australian Mobile Telecommunications Association's (AMTA) websites and they were then approached by phone calls and emails. Twenty two key informant representatives of these organizations agreed to be interviewed. They were considered experts within their organizations on mobile services topics. One of the challenges in selecting participants was to have at least one representative from each stakeholder category so that the views of each stakeholder are taken into account. Some of the organizations that were interviewed carry out multiple roles in the industry as shown in Table 2.

Figure 1 and Table 2 reveal that the sample of the interviews is generally representative of the Australian mobile telecommunications industry. Admittedly, we were unable to interview representatives from MVNOs. However, because the carriers occupy most of the market share in Australia lack of MVNO representation, though a limitation of the study, it cannot invalidate the qualitative findings reported in this paper (Informa, 2005a). The major carriers, their respective market shares, their growth prospects have been summarized in the Appendix to this paper. Also the Appendix includes the number of subscribers in Australia, the nature of the services used, the balance between prepaid and postpaid services and their growth prospects until 2010.

The interviews were focused on the mobile industry stakeholders and their roles; their perceptions on existing regulation and its impact on their operations; lessons learnt and experiences, and future directions. The contents of the interview transcripts were analyzed thematically. Codes were developed which provided the basis for cross-case analysis and helped identify and analyze emerging patterns of themes (Carson *et al*., 2001; Miles & Huberman, 1994). The themes were subsequently categorized and validated with existing literature before representative quotations were shortlisted from the interview for illustration purposes.

We believe that construct validity has been adequately addressed. First, multiple sources of information were used (Yin, 1994). While interviews constitute the primary source of information, some of the informants provided supporting secondary data which is comprised of archival white papers, professional journal articles and press releases, and web resources. In addition, the investigators themselves identified additional supporting documentation including materials located at the websites of the informants' organizations and industry publications. These secondary data were used for verifying and triangulating the findings of the interviews. Second, the informants belong to different categories of the mobile telecommunications industry. This ensured that different perspectives were provided which constitutes an important type of triangulation of qualitative information sources by preventing biased opinions (Choudhrie, Papazafeiropoulou, & Lee, 2003; Patton, 1990). Third, two investigators conducted and analyzed all interviews (Denzin, 1989). This kind of triangulation reduces the potential bias which is commonly cited as a limitation of interviews (Yin, 1994). Finally, the chain of evidence, tracing the conclusions to the interview summary and to the interview transcripts was also maintained. According to Yin (1994), these enhance the construct validity as well as the reliability of the research, thereby boosting its overall quality. Clearly, the study reported in this paper is based on the Australian context, and therefore, we accept its external validity cannot be ensured. Consequently, our findings may not be readily generalizable beyond this study (Shanks, Rouse, & Arnott, 1993). To ensure generalizability, further research is required, both in Australia and in other contexts.

ANALYSIS

In this section, we use the ideas presented in the work of (Kijl *et al*., 2005; Killström *et al*., 2006)) as a basis to analyze our findings. Specifically, we examine five major regulatory issues, namely, consumer and intellectual property protection, market and resources access, and environment protection. These and the manner in which they operate in the Australian context are discussed in detail in the sections that follow and they form the backbone of a regulatory framework which we propose later in the paper.

Consumer Protection

Privacy is the right of individuals to be left alone and to exercise control over one's personal information. Privacy includes information privacy which represents the way in which organizations handle an individual's personal information (Minch, 2004; Newman & Bach, 2004; Ng-Kruelle *et al*., 2001). Privacy has been one of the core issues in the internet environment and it has similar effect in the mobile context (Minch, 2004; Newman & Bach, 2004; Ng-Kruelle *et al*., 2001). All interviewees were consistent in indicating that their organizations have strict policy provisions in place for safeguarding the privacy of consumer information with some even having implemented encryption mechanisms. These provisions were consistent and compliant with Australia's Privacy Act of which there was strong awareness amongst all interviewees (Ng-Kruelle *et al*., 2001).

Privacy becomes particularly important with context-aware mobile services whereby information related to users' context is collected. Ineffective regulatory conditions can be exploited by unethical market players that interface with end-users in an attempt to create unfair strategic advantage by using the end-users' personal data (Ubacht, 2004). Consequently, the end-users' perceived credibility and trust in the mobile environment can be adversely affected, particularly, if undesirable opportunistic behaviors occur (Rao & Troshani, 2007; Troshani & Rao Hill, 2008). Specific legislation and regulation is, therefore, required for safeguarding end-users' rights of privacy (Figge, 2004; Minch, 2004).

Interviewees also indicated that privacy regulation in the mobile industry has also considerably restricted the range of mobile services that are being offered in the Australian market. For example, Interviewee 7 describes Australian carriers to be "absolutely paranoid about privacy" which has affected the uptake of certain mobile services, including location-based services. In general, regulation can potentially be either a driver or a "big barrier" (Interviewee 7) for the diffusion of mobile services as captured in the following statement:

> … *Australia's prime supervisions of privacy … have a definite influence on the way that the mobile community has formed, because of what you can do in terms of marketing on a mobile phone, which is different in Europe. So of course they have services there that we don't have here because their regulatory environment enables that.* (Interviewee 1)

Protection of Minors There was broad agreement amongst interviewees that policies and legislation should be in place for regulating the access of offensive and harmful mobile services and content including adult content, inappropriate gaming, and gambling services from minors. Although certain types of mobile services (e.g. adult and quasi-adult services), have been driving demand and generating substantial revenues for many supply-side stakeholders, the mobile industry is making efforts to provide these services responsibly into the Australian market (ACMA, 2006). In response to these concerns, the government regulator and industry have been collaborating to design measures for reviewing the applicability of regulation of content and services delivered to mobile phones. The first step towards the development of these measures is classification of mobile content and services (ACMA, 2006).

Further, Australian carriers have taken initiatives for implementing robust access controls and age verification processes for restricted content in order to eliminate the possibility of minors accessing inappropriate services. In general, all stakeholders have been proactive in promoting the safety of minors in the mobile space (AMTA, 2006). The efforts of the Australian mobile industry have culminated with the new Restricted Access Systems Declaration (RASD) which places obligations on all supply-side stakeholders to subject offerings of mobile services including mobile portals and premium rate SMS/MMS services to age-restricted access rules (ACMA, 2007a, 2007b).

Contractual Relationships refers to protecting the contractual bilateral relationships between supply- and demand-side stakeholders. This includes protecting end-users in their relationships with mobile network operators with mobile retail tariff regulation, mobile number portability, spamming, and product and billing disclosure (ACMA, 2006). Interviewees were consistent in their expectations that the government regulator should collaborate with industry to regulate these issues and create an industry code of conduct.

Interviewee 11 argues that regulating these aspects is particularly important in the mobile industry which may be particularly susceptible to fraud. Interviewee 12 illustrates:

[some content providers] are running ring tones and screen savers at the moment. They're targeting the youth market. They send an SMS saying do you want a free ring tone. If you say yes, you've found that you haven't just signed up for a free ring tone, you've signed up for a monthly subscription for twelve months for a ring tone at [A]$7 a month. This just turns up on your bill. This is destroying [consumer] trust and confidence. (Interviewee 12)

The devious activities of unscrupulous players in the mobile industry have a strong potential of creating mistrust, and in turn, seriously damage the reputation of legitimate supply-side stakeholders (Woolfson, 2005). A trusting environment is extremely important to exist in the industry for the mobile services to be widely adopted (Troshani & Rao Hill, 2008).

According to the Australian Telecommunications Industry Ombudsman (TIO), mobile complaints represent approximately 43% of total complaints (ACMA, 2006; Informa, 2005b). In response to a dramatic increase in complaints about SMS-based premium services, such as ringtones and horoscope downloads, particularly, from young adults, and in collaboration with the regulator, Australian content providers have implemented the Mobile Premium Services Scheme. Under this scheme, content providers are required to indicate clearly whether automatic subscription is involved with any purchase and all associated costs (ACMA, 2007b). In addition, the government regulator has collaborated with the industry to issue guidelines governing mobile commerce practices which expect both mobile operators' and content providers' compliance. These guidelines aim to protect consumers from dubious marketing practices, and educate mobile content providers about their responsibilities concerning the disclosure of information on pricing, consumer contracts, and terms and conditions (Rao & Troshani, 2006).

Intellectual Property Protection

Refers the use of copyright, patent, trademark, digital rights management (DRM) and electronic licensing schemes for protecting intellectual property rights associated with mobile services provisioning. Effectively deployed DRM solutions create additional ways to generate revenue and ensure that both consumers and content developers trust the mobile channel. A key issue facing creators and distributors of mobile content is protection from piracy and unauthorized use (Craig & Graham, 2003).

In Australia, mobile content providers depend on the widespread adoption of DRM technologies and on the Intellectual Property Rights (IPR) regime (Liu, Safavi-Naini, & Sheppard, 2003; Rao & Troshani, 2006, 2007a, 2007b). All interviewees, content providers and application developers in particular, did recognize the threat that piracy poses to ownership rights. Although all expressed concerns about this phenomenon, many had accepted it as "fact of life" (Interviewee 22). Interviewees were, however, consistent in suggesting that intellectual property protection or lack thereof does influence the decision whether to use a channel for distributing their content or services. Most interviewees indicated that precautionary measures are undertaken independently intended at curtailing intellectual property violations. These include screening the trustworthiness of distribution channel partners and using technical solutions to curtail piracy.

Market Access

Market access concerns with accessing markets with appropriate mobile frequencies and obtaining the required operational licenses. It also includes competition and antitrust legislation which is aimed at ensuring that competition is not distorted and, consequently, consumers receive quality mobile services. The mobile market is unique in the telecommunications domain. This is because licenses are technically and politically limited (Ubacht, 2004). Although licensing is meant to occur in a competitive manner, new carriers are not likely to be on equal footing with incumbent ones as the latter have advantageous starting positions due to prior well established presence and market power often through vertical integration with other industry players (Ubacht, 2004). Consequently, asymmetric market positions of the network operator players may become a source of regulatory concerns. Many interviewees argue that, currently, Australian supply-side stakeholders do not have equitable access to the market of mobile content and services, and that the government regulator should interfere by creating "a level playing field" (Interviewee 1) for all. Two main issues were raised:

First, government favouritism for one of the mobile network operators has the potential to create monopolistic behaviour in the mobile market. The interviewees representing the carriers, in particular, argued that there should be market access regulation:

The difficulty the government [regulator] has at the moment of their own design is that the government is the major shareholder in the largest telco in Australia (Interviewee 6)

The carrier interviewees were consistent in arguing that this position of power of the government-owned operator has affected competition for market access:

[The] key for companies such as ours is to make sure that where there's predatory market behavior, the government and the regulator make sure that it's not going to the detriment of competition. So if you look at - bundling is a big one for us and a lot of the smaller telcos, where companies like XXXXX can bundle, say you can bundle your home, your mobile, your business, your [cable/satellite TV], you can put them altogether and provide a discount. That's monopolistic behavior which makes it very difficult for competition. (Interviewee 6)

While there is no evidence that bundling strategies, as suggested above, have wielded substantial market power, strategies of this nature do have the potential of creating anti-competitive market domination of the larger operators over the smaller ones. Consequently, the Australian regulator will continue monitoring the market for further developments (ACMA, 2007b).

Second, the non-operator supply-side stakeholders were generally consistent in arguing that there is a power imbalance between carriers and the remaining supply-side stakeholders. This was of concern to many because the fact that carriers control billing and customer relationships gives them “the upper hand” (Interviewee 2) in controlling revenue sharing schemes with content providers. In fact, these schemes were perceived by many to be “unfair”, thus, creating a “huge barrier” (Interviewee 2) for content providers. Nevertheless, the majority of the interviewees were consistent in indicating that revenue sharing should be dictated by competitive market dynamics rather than by regulatory intervention.

Resources Access

Resources access includes knowledge development and deployment as well as subsidies provided by the government. These are discussed next.

Knowledge Development and Deployment The creation of technical and business knowledge underlying the development of mobile content and services is essential for the success of emerging industries. Currently, many content providers such as digital media organizations, artists, musicians and film makers have exhibited a huge interest for distributing their content via the mobile channel, however, the knowledge concerning the ways that mobile content can be adequately formatted and commercialized for mobiles is limited. Funding research or coordinating taskforces that build this knowledge are two possible options for mobile industry stakeholders (Choudhrie, Papazafeiropoulou, & Lee, 2003; Damsgaard & Lyytinen, 2001; King *et al.*, 1994). Currently, the content providers’ industry association in collaboration with government and a few key players in the industry have jointly undertaken active knowledge building steps, such as research, for developing this body of knowledge.

Once built, development knowledge and technical know-how needs to be deployed and this is important not only for building awareness amongst stakeholders, but also for showing them how business models operate. Interviewee 8 argues that many:

…[supply-side stakeholders] do not understand the technology” and “there are very few points of contact where they [stakeholders] can be educated. (Interviewee 8)

Therefore, industry associations should become more proactive in undertaking additional knowledge deployment measures including education and training. These measures can help supply-side stakeholders acquire the necessary knowledge and learn about the ways that they can format mobile content and services for mobile media, and distribute to end-users. As part of knowledge deployment the interviewees emphasized the need for publicizing success stories of exemplar local supply-side stakeholders in the local media and industry newsletters. This, they argue, is likely to create a conducive environment for the emergence of a social learning phenomenon based on the “law of imitation” (Hamblin, Miller, & Saxton, 1979). Publicized success stories set examples and could help new entrants in the industry learn about critical success factors and lessons learnt from past experiences. Interviewees were consistent in indicating that although there have been several success stories of local content providers in Australia, limited few have been widely publicized in local media.

Subsidies Often governments, industry associations, and other powerful players in the market may provide subsidies to players in emerging industries (Choudhrie, Papazafeiropoulou, & Lee, 2003; Damsgaard & Lyytinen, 2001; King *et al.*, 1994; Muzzi & Kautz, 2004). All interviewees indicated that currently, in the Australian mobile industry subsidies have mainly been in the form of small grants to fund trips to mobile fairs, exhibitions, and industry conferences to help small stakeholders showcase their products and facilitate networking with key stakeholders. Also, industry associations such as AIMIA systematically provide free access to market intelligence (Mackay & Weidlich, 2007). However, many believe that this level of subsidy support is limited. Many stakeholders in the mobile industry, particularly, content providers and application developers, believe that they are in a disadvantaged position in Australia. First, they believe that there is a lack of risk-sharing in the Australian industry which hinders development. For example, developers of games and mobile TV episodes incur substantial development costs, but share revenues with carriers or aggregators (Interviewee 21). When mobile services fail to generate expected revenues, losses are borne solely by developers. Second, many interviewees indicated that current funding and incentive schemes are inadequate. They expect government and industry associations to help by funding innovative services, and research and development initiatives.

Environmental Protection

Environmental protection is concerned with how developments in the mobile industry affect the environment. The Australian mobile industry has undertaken a transparent, viable and sustainable mobile phone recycling campaign with the objective of preventing used mobile phones, accessories and batteries end up in landfill and as a result minimize environmental burden (AMTA, 2006). With increasing handset replacement rates (Informa, 2005a, 2005b), most interviewees support the mobile phone recycling campaign and hope that this initiative will start a bandwagon effect amongst all stakeholders.

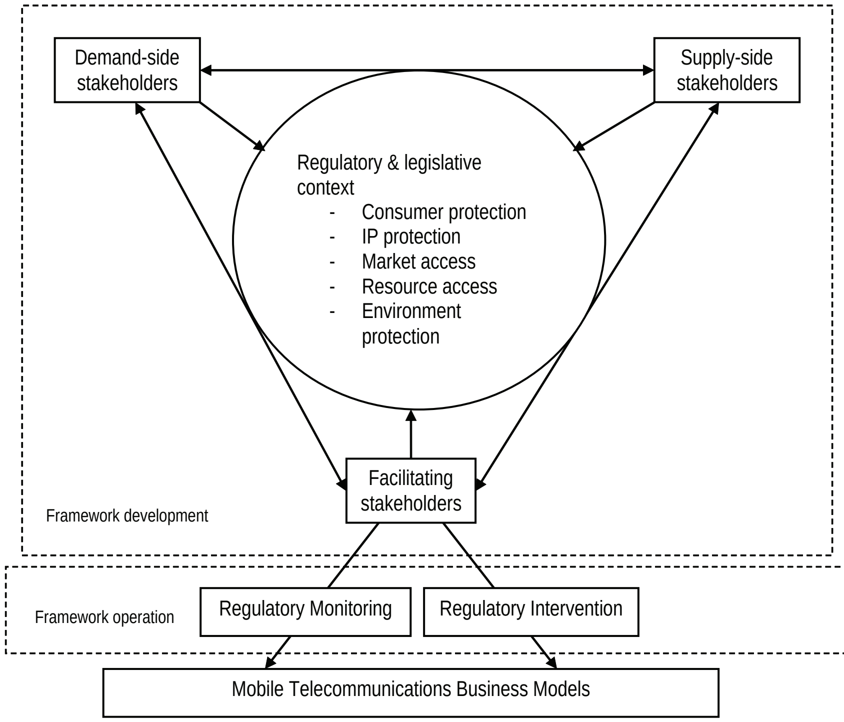
REGULATORY FRAMEWORK FOR MOBILE SERVICES

In general, regulation setting is implemented by organizations with legislative powers, such as regulatory bodies. These regulate the context in which products or services are developed, deployed, and used. For emerging industries such as the mobile services industry, regulatory authorities have to deal with a multitude of heterogeneous networked stakeholders. Further, as mobile technology operates in a dynamic environment and undergoes rapid changes, regulatory definitions become a moving target which implies that regulators should constantly acquire industry-specific knowledge over time (Tallberg *et al*., 2007; Ubacht, 2004). Consequently, regulation for the mobile industry becomes an extremely complex and involved undertaking for all stakeholders.

The Australian government policy for the mobile industry is characterized by a co-regulation approach which represents close collaboration with the industry in terms of a mixture of direct intervention through legislation, on the one hand, and complete self-regulation on the other. There is no direct regulation, but also there is no pure self-regulation. Regulatory bodies provide the industry with some parameters concerning the regulatory issues discussed in the previous section in which key problems are to be solved. It is the responsibility of the industry to work out the details that best suits mobile technology and related business models. The role of the regulator is thus to allow the industry to apply its own codes in the first instance and to monitor the effectiveness and enforcement of those codes (AMTA, 2005a).

The diagram in Figure 2 integrates the regulatory issues discussed previously with the notion of co-regulation to form our regulatory framework for the mobile industry in Australia. These issues are feasible for two reasons. First, they were presented by the interviewees and feature heavily in the literature as important prerequisites for the success of the mobile industry. Second, these issues have been implemented in various forms in other industries. These issues fit together in the framework in that they represent the interests of all stakeholders and are shaped by their joint

Figure 2. Co-regulation in the mobile industry in Australia



efforts. For example, while consumer protection is geared towards protecting end-users, market access protects smaller or weaker supply-side stakeholders against predatory practices of more powerful ones. As indicated in Figure 2, the consideration of these issues is made by involving the industry as a whole instead of individual stakeholders which is likely to result in balanced and sustainable outcomes in the long term.

Figure 2 also shows that the regulatory framework operates via compliance monitoring and intervention. First, monitoring may be implemented by establishing suitable reporting mechanisms. Second, intervention should only occur in cases of compliance violations or market failure. Regulatory issues included in our framework, while jointly developed by all stakeholders, they do represent different but equally important stakeholders. Consequently, these issues are generally similarly equal in importance and complementary. It follows that their interaction is likely to be frictionless. However, if conflicts do occur, facilitating stakeholders should intervene. In any case, the development and operation of the proposed regulatory framework should not interfere with or distort future market or industry developments.

With co-regulation, the mobile industry is empowered and takes responsibility for participating in the development of its own regulation. Three major benefits emerge with this approach: first, regulation costs are likely to be significantly reduced; second, compliance is likely to occur naturally, and therefore, regulation in itself is likely to be perceived to be less restrictive and onerous than in traditional regulation models (ACMA, 2006; AMTA, 2006); third, industry-driven co-regulation also has the advantage to ensure that it is likely to remain appropriate and be responsive to changing market conditions and capable of delivering timely and transparent outcomes. Taken together these advantages are likely to promote business activity, market integrity, and investor confidence in the Australian telecommunications industry and its international competitiveness.

CONCLUSION AND FUTURE RESEARCH

Regulating the mobile services industry is an extremely difficult task. There are many reasons for this, including the highly complex nature of the networks and stakeholder relationships required to provide mobile services as well as the constantly evolving underlying technologies. There are growing calls from both scholars and practitioners alike for further research in this area. In response to these calls we have conducted a study which draws both on existing literature and qualitative evidence and have proposed a regulatory framework for mobile services. We have first provided an overview of the Australian mobile services industry which was used as a setting for our investigation. We have subsequently discussed the research method followed by an elaboration of the proposed framework. We believe that this framework is the first of its kind, and, thus, it contributes to the existing body of knowledge which can be employed by both academics and practitioners alike.

The framework is comprised of five major components: first, consumer protection including privacy, protection of minors, contractual relationships, second, intellectual property protection, third, market access, fourth, resources access including knowledge development and deployment, and subsidies, and finally, environmental protection. We argue that these encompass the interests of all stakeholders operating mobile industry and given its dynamic and complex nature co-regulation is the most effective approach to minimize costs and enhance compliance.

Using our framework as a starting point, there are several directions for future research. First, the proposed framework may not be exhaustive. Thus, further studies encompassing both wealthy and low income countries that can enable generalizations and further validation of the proposed framework should be undertaken. Second, further work should focus on enforcement mechanisms of the components of the framework. Third, different countries represent different regulatory jurisdictions which can be diverse (Verikoukis *et al*., 2006). As various supply-side stakeholders (e.g. network operators) increasingly offer mobile services across national borders, further research is required for developing regulatory interfaces for transnational mobile markets. This is an important first step for providing regionally or internationally unified regulation, thereby, increasing the current limited investment possibilities that exist in fragmented markets. Fourth, mobile communication platforms, services, content and devices are converging and changing rapidly. The challenge for both academics and practitioners alike is to investigate whether and how convergent and dynamic mobile services can be regulated without succumbing to the supposition that these constitute new problems requiring new regulatory solutions. Therefore, the question if regulation can be 'technologically-neutral' requires further investigation.

REFERENCES

ACMA. (2006). *Telecommunications Performance Report 2004-05.* Sydney: Australian Communications and Media Authority (ACMA).

ACMA. (2007a). *New rules for age-restricted internet and mobile content* (ACMA Media Release 151/2007 - 21 December). Retrieved 11 January, 2008, from http://www.acma.gov.au/WEB/STANDARD/pc=PC_310907

ACMA. (2007b). *New restricted access arrangements: Restricted Access System Declaration 2007.* Retrieved 11 January, 2008, from http://www.acma.gov.au/WEB/STANDARD/pc=PC_310905

Alahuhta, P., Ahola, J., & Hakala, H. (2005). *Mobilising business applications: a survey about the opportunities and challenges of mobile business applications and services in Finland*. Helsinki: Tekes.

AMTA. (2005a). *AMTA Response to regulation taskforece isues paper 2005.* Australian Capital Territory: ACT: Australian Mobile Telecommunications Association.

AMTA. (2006). *Australian Mobile Telecommunications Association (AMTA) 2006 Annual Report.* Australian Capital Territory (ACT): Australian Mobile Telecommunications Association.

Bina, M., & Giaglis, G. M. (2005). *Exploring early usage patterns of mobile data services.* Paper presented at the International Conference on Mobile Business, Sydney, Australia.

Bouwman, H., Haaker, T., & De Vos, H. (2007). Mobile service bundles: the example of navigation services. *Electronic Markets, 17*(1), 20-28.

Buerkler, E. (2005). Dramatically different price developments for mobile services in India and Switzerland and their underlying drivers. *Info, 7*(2), 72-75.

Camponovo, G., & Pigneur, Y. (2003). *Business Model Analysis Applied to Mobile Business.* Paper presented at the 5th International Conference on Enterprise Information Systems (ICEIS 2003), Angers, France, 23-26 April.

Carson, D., Gilmore, A., Gronhaug, K., & Perry, C. (2001). *Qualitative Research in Marketing.* London: Sage.

Cassell, C., & Symon, G. (1994). *Qualitative Methods in Organisational Research*. London: Sage Publications.

Chen, L. (2008). A model of consumer acceptance of mobile payment. *International Journal of Mobile Communications, 6*(1), 32-52.

Choudhrie, J., Papazafeiropoulou, A., & Lee, H. (2003). A web of stakeholders and strategies: a case of broadband diffusion in South Korea. *Journal of Information Technology, 18*(4), 281-290.

Craig, C., & Graham, R. (2003). Rights management in the digital world. *Computer Law & Security Report, 19*(5), 356-362.

Damsgaard, J., & Lyytinen, K. (2001). The role of intermediating institutions in the diffusion of electronic data interchange (EDI): how industry associations intervened in Denmark, Finland, and Hong Kong. *The Information Society, 17*(3), 195-210.

Denzin, N. K. (1989). *The Research Act: A Theoretical Introduction to Sociological Methods* (3rd ed.). Englewood Cliffs, N. J.: Prentice Hall.

Figge, S. (2004). Situation-dependent services - a challenge for mobile operators. *Journal of Business Research, 57*(12), 1416-1422.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research Methods in the Social Sciences* (5th ed.). New York: St. Martin's Press.

Gelenbe, P. (2003). Will 'big brother' fears halt success of location-based mobile services. *Precision Marketing, 15*(15), 16-17.

Goggin, G., & Spurgeon, C. (2005). Mobile message services and communications policy. *Prometheus, 23*(2), 181-193.

Grzybowski, L. (2005). Regulation of mobile telephony across the European Union: an empirical analysis. *Journal of Regulatory Economics, 28*(1), 47-67.

Hamblin, R. L., Miller, J. L. L., & Saxton, D. E. (1979). Modeling use diffusion. *Social Forces, 57*(3), 799-811.

Han, S., Choi, S. H., Kim, B., & Chung, H. (2006). A quantitative analysis of the effects of a handset subsidy on consumer welfare. *Electronics and*

Telecommunications Research Institute (ETRI) Journal, 28(5), 621-630.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal, 3*(3), 118-126.

Hsieh, C., Jones, C., & Lin, B. (2008). The new business potential with mobile commerce. *International Journal of Mobile Communications, 6*(4), 436-455.

Informa. (2005a). *Mobile Content and Services: Worldwide Market Analysis & Strategic Outlook*. London: Informa Telecoms & Media.

Informa. (2005b). *Global Mobile Forecasts to 2010: Worldwide Market Analysis & Strategic Outlook 2005-2010* (5th ed.). London: Informa Telecoms & Media.

Informa. (2006a). Global Mobile: The International Business Newsletter of Mobile Communications Markets. *13*(6).

Jho, W. (2007). Liberalisation as a development strategy: network governance in the Korean mobile telecom market. *Governance: An International Journal of Policy, Administration, and Institutions, 20*(4), 633-654.

Kijl, B., Bouwman, H., Haaker, T., & Faber, E. (2005). *Dynamic Business Models in Mobile Service Value Networks: A Theoretical and Conceptual Framework (FRUX/D2.2-II).* The Netherlands: TU Delft.

Killström, U., Galli, L., Haaker, T., Immonen, O., Kijl, B., Pitkänen, O., Saarinen, P. J., & Virola, H. (2006). *Initial marketplace dynamics (incl. business models) analysis (IST-2004-511607 MobiLife).* Finland: Information Society Technologies.

King, J. L., Gurbaxani, V., Kraemer, K. L., McFarlan, K. L., Raman, K. S., & Yap, C. S. (1994). Institutional factors in information technology innovation. *Information Systems Research, 5*(2), 139-169.

Kumar, S. (2004). Mobile communications: global trends in the 21st century. *International Journal of Mobile Communications, 2*(1), 67-86.

Li, W., & McQueen, J. (2008). Barriers to mobile commerce adoption: an analysis framework for a country-level perspective. *International Journal of Mobile Communications, 6*(2), 231-257.

Liu, Q., Safavi-Naini, R., & Sheppard, N. P. (2003). Digital rights management for content distribution. In J. P. Montague & C. Steketee (Eds.), *Conferences in Research and Practice in Information Technology: Australasian Information Security Workshop 2003 (AISW2003))* (Vol. 21). Adelaide: Australian Computer Society.

MacInnes, I. (2005). Dynamic business model framework for emerging technologies. *International Journal of Services Technology and Management, 6*(1), 3-19.

Mackay, M. M., & Weidlich, O. (2007, March 2007). Australian Mobile Phone Lifestyle Index, 3rd Edition, Australian Interactive Media Industry Association (AIMIA), Mobile Content Industry Development Group (MCIDG). Retrieved 8 April, 2007, from http://www.aimia.com.au/mobile

Massey, A. P., Khatri, V., & Ramesh, V. (2005). *From the web to the wireless web: technology readiness and usability.* Paper presented at the 38th Hawaii International Conference on System Sciences, Hawaii.

Miles, M. B., & Huberman, A. M. (1994). *An Expanded Sourcebook: Qualitative Data Analysis.* Thousands Oaks: Sage.

Minch, R. P. (2004). *Privacy issues in location-aware mobile devices.* Paper presented at the Proceedings of the 37th Hawaii International Conference on System Sciences, January 5-8

Muzzi, C., & Kautz, K. (2004). Information and communication technologies diffusion in industrial districts. In J. Damsgaard & H. Z. Henriksen (Eds.), *Networked Information Technologies: Diffusion and Adoption* (pp. 19-38). Dordrecht: Kluwer Academic Publishers.

Mylonopoulos, N. A., & Sideris, I. A. (2006). Growth of value added mobile services under different scenarios of industry evolution. *Electronic Markets, 16*(1), 28-40.

Neuman, W. L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches* (4th ed.). Boston: Allyn and Bacon.

Newman, A. L., & Bach, D. (2004). Privacy and regulation in a digital age. In B. Preissl, H. Bouwman & C. Steinfield (Eds.), *E-Life After the Dot. Com Bust*. Heidelberg: Physica-Verlag.

Ng-Kruelle, G., Swatman, P. A., Rebne, D. S., & Hempe, J. F. (2001). The price of convenience: Privacy and mobile commerce. *Quarterly Journal of Electronic Commerce, 3*(4), 273-285.

Oh, S., Yang, S., Kurnia, S., Lee, H., Mackay, M. M., & O'Doherty, K. (2008). The characteristics of mobile data users in Australia. *International Journal of Mobile Communications, 6*(2), 217-230.

Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.). London: Sage Publications.

Pitkänen, O. (2006). *Legal and regulation framework specification: competence within mobile families and ad-hoc communities (IST-2004-511607 MobiLife)*. Finland: Information Society Technologies.

Rao, S., & Troshani, I. (2006). *AMC-ER Preliminary Research Report: Understanding Market and Mobile Content Selection for Distribution*. Adelaide, South Australia: Australian Mobile Content – Export Research Initiative, m.Net Corporation Ltd.

Rao, S., & Troshani, I. (2007). A conceptual framework and propositions for the acceptance of mobile services. *Journal of Theoretical and Applied Electronic Commerce Research 2*(2), 61-73.

Rao, S., & Troshani, I. (2007a). *AMC-ER Challenges and Rewards: Experiences of Australian Content Exporters*. Adelaide, South Australia: Australian Mobile Content – Export Research Initiative, m.Net Corporation Ltd.

Rao, S., & Troshani, I. (2007b). *AMC-ER Initial Selection of Content, Markets, and Distribution Channels for the Australian Mobile Content – Export Research Initiative*. Adelaide, South Australia: m.Net Corporation Ltd.

Sangwan, S., & Pau, L.-F. (2005). Diffusion of mobile terminals in China. *European Management Journal, 23*(6), 674-681.

Shanks, G., Rouse, A., & Arnott, D. (1993). *A review of approaches to research and scholarship in information systems*. Paper presented at the Fourth Australian Conference in Information Systems, Brisbane, Queensland.

Smith, A. M., & Fischbacher, M. (2005). New service development: a stakeholder perspective. *European Journal of Marketing, 39*(9/10), 1025-1048.

Spiller, P. T., & Zelner, B. A. (1997). Product complementarities, capabilities and governance: a dynamic transaction cost perspective. *Industrial and Corporate Change, 6*(3), 561-594.

Steinbock, D. (2005). *The Mobile Revolution: The Making of Mobile Services Worldwide*. London: Kogan Page.

Tallberg, M., Hämmäinen, H., Töyli, J., Kamppari, S., & Kivi, A. (2007). Impacts of handset bundling on mobile data usage: the case of Finland. *Telecommunications Policy, 31*(10/11), 648-659.

Tongia, R. (2007). Connectivity in emerging regions: the need for improved technology and business models. *IEEE Communications Magazine, 45*(1), 96-103.

Troshani, I., & Rao Hill, S. (2008). A Proposed Framework for Mobile Services Adoption: A Review of Existing Theories, Extensions, and Future Research Directions. In G. C. Karmakar & L. S. Dooley (Eds.), *Mobile Multimedia Communications: Concepts, Applications and Challenges* (pp. 85-108). Hershey, PA, USA: IGI Global.

Troshani, I., & Rao, S. (2007a). The diffusion of mobile services in Australia: an evaluation using stakeholder and transaction cost economics theories. *IADIS International Journal of WWW/Internet, 5*(2), 40-57.

Ubacht, J. (2004). Regulatory practice in telecommunications markets: how to grasp the 'real world of public action'? *Communications & Strategies*, (54, 2nd Quarter), 219-242.

Van De Kar, E. A. M. (2004). *Designing Mobile Information Services: An Approach for Organisations in a Value Network*. Unpublished Doctoral Dissertation, Delft University of Technology, Delft.

Verikoukis, C., Konstas, I., Anastasiadou, D., & Angelidis, P. (2006). Overview on telecommunications regulation framework in south-eastern Europe. *Telecommunication Systems, 32*(2/3), 209-221.

Woolfson, M. (2005). Can self-regulation protect the mobile industry from crooks? *New Media Age, 17/03/2005*, 16.

Yin, R. K. (1994). *Case Study Research: Design and Methods*. Beverley Hills: Sage.

Yoo, Y., Lyytinen, K., & Yang, H. (2005). The role of standards in innovation and diffusion of broadband mobile services : The case of South Korea. *Journal of Strategic Information Systems, 14*(3), 323-353.

KEY TERMS

Contractual Relationships: Contractual Relationships refer to protecting the contractual bilateral relationships between supply- and demand-side stakeholders. This includes protecting end-users in their relationships with mobile network operators with mobile retail tariff regulation, mobile number portability, and product and billing disclosure.

Co-Regulation: Co-regulation represents a regulation approach by way of which government regulations collaborate closely with the industry in terms of a mixture of direct intervention through legislation on the one hand, and complete self-regulation on the other.

Environmental Protection: Environmental protection is concerned with how developments in the mobile industry affect the environment.

Intellectual Property Protection: Intellectual Property Protection refers the use of copyright, patent, trademark, digital rights management (DRM) and electronic licensing schemes for protecting intellectual property rights associated with mobile services provisioning.

Market Access: Market Access is concerned with accessing markets with appropriate mobile frequencies and obtaining the required operational licenses. It also includes competition and anti-trust legislation which is aimed at ensuring that competition is not distorted and, as a consequence, consumers receive quality mobile services.

Mobile Services: A mobile service is an activity or series of intangible activities that occur when mobile consumers interact with systems or service provider employees with the support of a mobile telecommunications network (Bouwman, Haaker, & De Vos, 2007; Van De Kar, 2004). Examples of mobile services include mobile e-mail, SMS/MMS services, content downloads, mobile ticket reservations, and mobile TV.

Privacy: Privacy is the right of individuals to be left alone and to exercise control over one's personal information. Privacy includes information privacy which represents the way in which organizations handle an individual's personal information.

Regulatory Regime: A regulatory regime represents any type of authority which can operate at either industrial, national or international levels and which can influence, direct, limit or prohibit any activity undertaken by any stakeholder operating in the mobile telecommunications industry (Yoo, Lyytinen, & Yang, 2005).

APPENDIX

Table 1a. Australian Mobile Market to 2010 (Source: Informa, Global Mobile Forecasts to 2010)

	2006	2007	2008	2009	2010
Total Subscriptions (000)	19,964	20,699	21,294	21,781	22,184
Growth Subscriptions (%)	5.10	3.68	2.87	2.29	1.85
Penetration (%)	98.11	100.89	102.96	104.49	105.62
Total Revenues (US$ million)	8,885	9,138	9,195	9,075	8,869
Growth Revenues (%)	5.04	4.59	4.33	3.90	3.81
Revenues (US$ million)					
Voice	7,627	7,650	7,474	7,121	6,673
Data	1,258	1,488	1,722	1,954	2,196
Revenue share (%)					
Voice	85.84	83.72	81.28	78.47	75.24
Data	14.16	16.28	18.28	21.53	24.76
Prepaid					
Total prepaid subscriptions (000)	8,739	9,173	9,524	9,811	10,049
Prepaid as % of total subscriptions	44	44	45	45	45
Subscriptions by operator (000)					
Telstra Mobile	8,994	9,265	9,503	9,698	9,859
SingTel Optus	6,859	6,942	7,061	7,158	7,239
Vodafone	2,707	2,883	3,002	3,099	3,180
Hutchison (Orange and 3 Australia	1,404	1,609	1,728	1,826	1,906
Operator market share (%)					
Telstra Mobile	45	45	45	45	44
SingTel Optus	34	34	33	33	33
Vodafone	14	14	14	14	14
Hutchison (Organge & 3 Australia)	7	8	8	8	9

Section II
Telecommunications Industry

Chapter VII
Analyzing the Disruptive Potential in the Telecommunications Industry

Stefan Hüsig
University of Regensburg, Germany

ABSTRACT

This chapter covers an important issue in the area of telecommunications planning and technology management: The ex ante analysis of potential disruptive technologies. Due to the convergence of IT and telecommunications, an ever-growing onslaught of emerging technologies and new entrants with new business models are starting to eat up the incumbent's revenues, profits, and market shares. The theory of disruptive technology (Christensen, 1997; Christensen & Raynor, 2003) helps managers, policy-makers and analysts to analyze emerging technologies, new business models, and new entrants in order to be prepared in advance, providing appropriate ways to react in a timely fashion to innovation-based opportunities or threats. In this chapter the theory of disruptive technology, the concept of disruptive potential and a method for applying this concept in a telecommunications planning and technology management context is presented. Finally, some examples of potential disruptive technologies in telecommunication which were analyzed ex ante are introduced with specific emphasis placed on the WLAN-technology.

INTRODUCTION

The telecommunications industry has been undergoing a radical transformation since the beginning of the new century, creating new opportunities and new challenges for infrastructure and service providers (Christensen, Anthony, & Roth, 2004; Lia & Whalleyb, 2002). The formerly established integrated value chain is increasingly being deconstructed, industry borders are becoming increasingly blurred due to convergence between the Internet and telecommunications industry is being radically restructured with the entry of new players (Bores, Saurina, & Torres, 2003; Lia & Whalleyb, 2002). Unexpectedly, discontinuous technological developments in various areas of the industry and increasing market turbulence have added new dimensions to an already complex scenario. Many traditional business models, as well as related frameworks, tools, and methods have become obsolete. Therefore, the development of new conceptual frameworks for understanding these changes in the telecommunications and related industries; and the creation of new tools and methods for identifying opportunities and threats, and making new strategies is becoming a central task for academia. In the new environment of uncertainty and ambiguity managers, policy-makers, researchers and analysts are in need of improved and new theoretical based frameworks to gather and interpret information in another way to finally recognize the opportunities and threats emerging from discontinuities faster and more effectively.

So far the industries' track record in innovation is quite mixed with 3G and the mobile Internet, Multimedia Messaging Service (MMS) or the attempts to transfer i-Mode to other contexts than Japan (Christensen et al., 2004; Hüsig et al., 2005; Lyytinen & King, 2002; Rao, Angelov, & Nov, 2006). Quite frequently innovations from outside the inner domain of the telecommunication industry are introduced by outsiders like Skype with Voice over Internet Protocol (VoIP) or Cisco Systems with their IP-routers which took and take away customers and revenues from the traditional incumbents in telecommunications (Christensen et al., 2004; Rao et al., 2006). Some traditional infrastructure suppliers in the telecommunication industry like Siemens or IBM already went almost completely out of this business with their circuit-switched technologies which were disrupted away by Cisco Systems – a new entrant with a new technology which the incumbents never managed to adopt effectively (Christensen et al., 2004). It seems that managers and analysts in telecommunications need better tools and methods to effectively identify and counteract disruptions like these. The theory of disruptive technology (Christensen, 1997; Christensen & Raynor, 2003) helps managers, policy-makers and analysts to analyze emerging technologies, new business models, and new entrants in order to be prepared in advance, providing appropriate ways to react in a timely fashion to innovation-based opportunities or threats. However, not many scholars have tried to apply this theory to the field of telecommunications (Christensen et al., 2004; Hüsig et al., 2005; Osterwalder, Ondrus, & Pigneur, 2005; Rao et al., 2006). Therefore, a more comprehensive and easily applicable method for identifying disruptive technology threats from an ex ante perspective had to be developed further, and is elaborated here. For managers, policy-makers and analysts in telecommunications a more applied version of the theory of disruptive technology in form of recently developed tool provides a better guideline for this important issue in the area of telecommunications planning and technology management.

In this chapter, we present first the disruptive technology theory, then the concept of disruptive potential, and a method of applying this concept in a telecommunications planning and technology management context. Finally, some examples of potential disruptive technologies in telecommunication which were analyzed ex ante are introduced with specific emphasis placed on the WLAN-technology.

THE CONCEPT OF DISRUPTIVE POTENTIAL

The Disruptive Technology Theory

In order to grasp the concept of disruptive potential it is essential to understand the theoretical underpinning first: The disruptive technology theory. The theory of disruptive technology or disruptive innovation originally came from Christensen's research into the rigid disk drive industry (Christensen, 1993; 1997). His research focused on the question of why and under which circumstances flourishing incumbents in a given industry could be successfully attacked by new entrants using new technologies. He identified a specific set of such technologies with distinct characteristics, which he termed "disruptive" technologies. Accordingly, these disruptive technologies are defined as technologies which disrupt an established trajectory of performance improvement downwards, or redefine what performance means (Christensen & Bower, 1996).

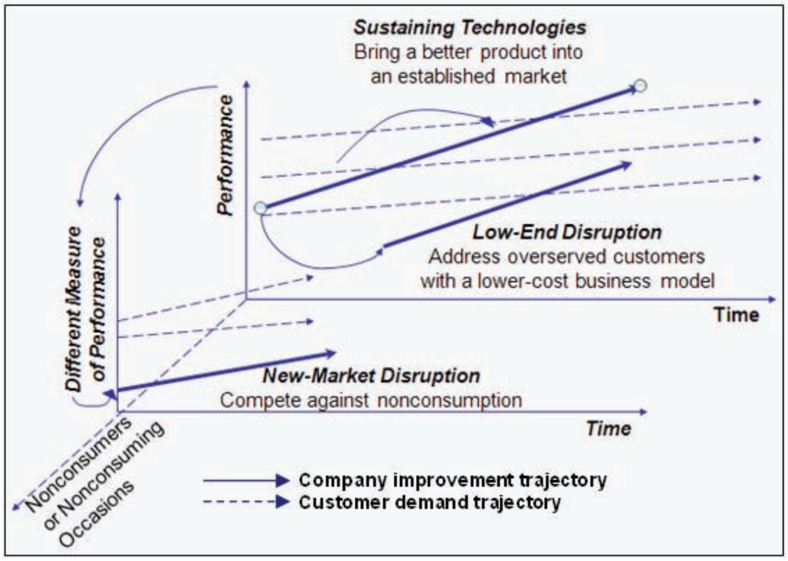
They stand in contrast to sustaining technologies, which are technological changes that have a sustaining impact on an established trajectory of performance improvement and give the known customers more and better versions of what they have valued up to now. These sustaining technologies lead to innovations that allow the firms to provide better and more profitable products to their existing customers (Christensen et al., 2004). Christensen's findings indicated that incumbent firms almost always ultimately master these sustaining innovations but are often unable to cope with disruptive ones. Even though disruptive technologies initially underperform established ones in serving the mainstream market, they eventually displace the established technologies (Danneels, 2004)

Christensen et al. (2004) differentiate two distinct types of disruptive opportunity: The "new-market disruption", in which a disruptive technology does not compete with other existing products but instead competes against non-consumption and therefore establishes an entirely new market, or the "low-end disruption" where a disruptive technology can be used to attack an existing value network from the low end by deploying a business model that profitably serves the less demanding customers whom incumbent market leaders are happy to shed as they themselves concentrate on more attractive customers. Nevertheless, in the long run the initially inferior disruptive technologies from new markets or the low end typically introduced by new entrants become good enough to meet the demands of the mainstream customers of the incumbents, so that the incumbents are forced toward higher margin customer groups, or out of the market entirely. This process of disruptive and sustaining technological change is best illustrated by Figure 1 below.

It is important to understand that Christensen defines the term technology quite broadly as "the processes by which an organisation transforms labour, capital, materials, and information into products and services" (Christensen, 1997; xiii). His process-based concept of technology extends beyond the engineering and manufacturing functions of the firm to also include business processes. Changes in these processes are innovations. Therefore, in his more recent publications the term "disruptive technology" (broadly understood) is now exchanged by "disruptive innovation" (Christensen et al., 2004; Christensen & Raynor, 2003). This means that Christensen's concept of technology is not limited to physical products but is explicitly appropriate for a service-based industry like telecommunications.

A disruptive innovation is typically financially unattractive for the leading incumbent to pursue, relative to its profit model and relative to other investments competing for the organization's resources. This translates to an advantage for entrants which operate with an alternative business model and do not have similar trade-offs between sustaining and disruptive technologies, since they would focus on the disruptive technology in a market segment which is new or unattractive to the incumbent. Therefore, disruptiveness is not an absolute phenomenon but a relative one which can only be measured relative to the business model of another firm. In other words, an innovation that is disruptive relative to the business model of one firm can be sustaining compared to the business model of another (Christensen et al., 2004).

Figure 1. The process of disruptive and sustaining technological change in a given value network (Christensen et al., 2004; Christensen & Raynor, 2003)



The example of Voice over Internet Protocol (VoIP) can illustrate the relativity of the same technology used in sustaining or disruptive ways: VoIP could be used in a sustaining way to lower costs and add new high-value services by the incumbents without altering much of their traditional business model. On the other hand, entrants like Skype used VoIP to provide lower quality voice services for free in their network to enter the value network of the traditional telecom incumbents from the low end (Christensen et al., 2004; Rao et al., 2006). Providing lower quality services for free was not in the business model of traditional incumbents therefore they had initially little reason to change. Used with Skype's business model, VoIP had more disruptive potential than in combination with the traditional model of the incumbents (Osterwalder et al., 2005). This case shows that almost the same technology can lead to a sustaining and disruptive outcome. Therefore the context, especially the value network and the specific business models have to be taken into account.

The explanation for the incumbents' behavior can be seen predominantly in the power of the established customers in the mainstream market (Christensen & Bower, 1996). According to Christensen, the inability to change strategy at the right time and allocate sufficient resources to the new technology for the emerging market results in the interaction between distinct circumstances in the internal resource allocation process of the firm (Bower, 1970; Burgelman, 1983a, 1983b). This effect is also suggested by resource dependence theory as developed by Pfeffer and Salancik (1978). According to this theory, the strategies firms can choose are strongly restricted by the interests of their existing customers and investors, who provide the resources necessary to survive. Therefore, established firms allocate their resources to sustaining technologies that address the interests of their existing customers rather than to disruptive technologies about which the customers and markets are highly uncertain, and which are initially structurally unattractive (Christensen, 1997; Christensen & Bower, 1996). The failure of the incumbent firms concerning disruptive technologies is based not on a deficiency in technological competence but on the firms' inability to change their strategy towards new value networks when there is little or no impetus from their existing and most valuable customers. The contrasting capabilities and asymmetrical motivations of incumbents and entrants in the face of disruptive technological change are based on different sets of resources, organizational processes, and values embedded in the organization's culture which are formed over time by their surrounding value networks (Christensen, 1997; Christensen et al., 2004; Christensen & Raynor, 2003; Christensen & Rosenbloom, 1995; Overdorf & Christensen, 2000).

Critique and Advancement of Christensen's Disruptive Technology Theory

With the growing popularity of Christensen's disruptive technology theory, the amount of criticism and extensions in the academic literature also increased. However, one must be careful when reading many of these papers, even if they have a scholarly source. Due to the complex argument of Christensen's theory, not all his critics fully grasp the entire complexity of the reasoning and come to misleading results. Nevertheless, it is also important to be aware of the limits of the theoretical framework in order to apply the following method successfully in a real life setting. Therefore, an overview of the most important and valid criticisms and advancements are briefly presented here:

- The theory still lacks terminological clarity and has methodological as well as formal deficits (Adner, 2002; Danneels, 2004; Govindarajan & Kopalle, 2006a, 2006b; Hüsig et al., 2005)
- The influence of institutional factors is still underdeveloped but might be important for the application and explanatory power of the theory especially in non-US contexts (Chesbrough, 1999, 2003; Hüsig, Dowling, & Hipp, 2007).
- The explanatory power via the resource allocation process and the resource dependence theory was questioned as the effects of threat perception might lead to different rigidities and resource commitments inside the incumbents as reactions to potentially disruptive technologies (Gilbert, 2002, 2005; Gilbert & Bower, 2002).
- The influence of specific industry and technological contexts such as standards, lock-in-effects and network externalities on the disruption process is still not fully understood nor has it been researched in depth (Adner & Levinthal, 2001; Hüsig et al., 2005).
- The explanatory power of previously marginal product attributes, which, Christensen suggests, explain the adoption of the disruptive technology in the mainstream market, has been attacked by empirical reconsideration of the original data, suggesting that the role of the lower unit price of disruptive technologies when invading the established market is a prominent force to explain and forecast the phenomenon (Adner, 2002).

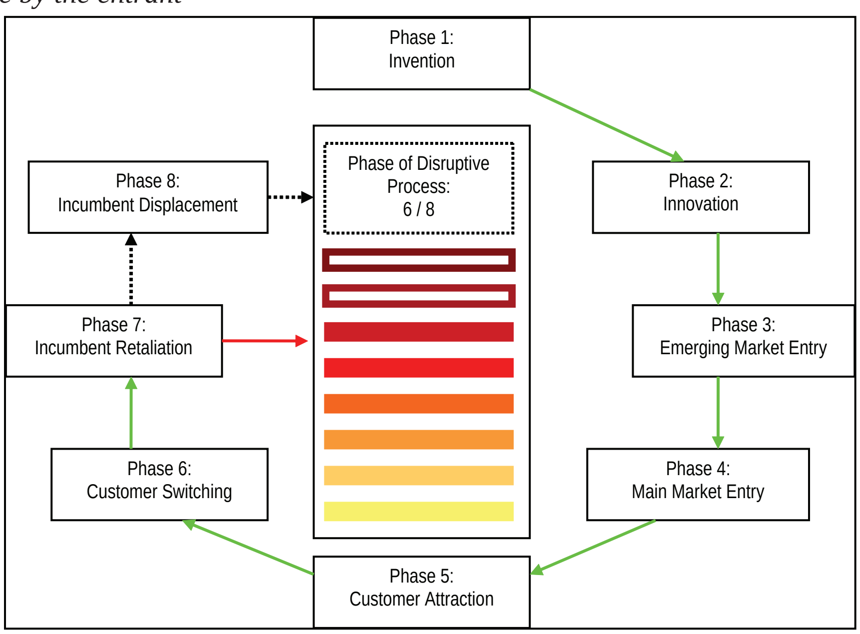
Some of these aspects could make it quite difficult to analyze the telecommunications sector using this theory as network effects, regulation, and scarce resources (e.g. frequencies) often play an important role in the adoption and diffusion of new services and technologies (Hüsig et al., 2005; Mannings & Cosier, 2001; Schoder, 2000). However, recent research in the telecommunications sector using the disruptive innovation framework shows the feasibility of the theory in this industry (Christensen et al., 2004; Hüsig et al., 2007; 2005). It seems that the incumbents in highly interdependent, networked industries provide better avenues for co-option (Christensen et al., 2004; Hüsig et al., 2005). In the case of telecommunications incumbents are often more motivated to fight the attack of an entrant building or acquiring the potential disruptive innovations themselves. The exit from low-end markets in scale intensive and high fix costs industries like telecommunications might be less attractive than in other industries where attractive high end markets also provide higher profits and margins for the ceding incumbents. Nevertheless, also here the theory of disruption helps to explain and predict under which circumstances the likelihood for co-option or ceding of the incumbents would be higher or lower.

Analysis of Emerging Technologies Using the Concept of Disruption

In a real life context, ex post reasoning is of little value to practitioners. They are more interested in an ex ante perspective in order to identify disruptive technological threats or opportunities to be able to formulate and implement a timely strategy. The concept of disruptive potential addresses this problem (Hüsig et al., 2005; Rafii & Kampas, 2002). Disruptive potential is a concept and measure for the ex ante analysis of the potential of emerging technologies for reaching the end of the disruption process successfully. The degree of disruptive potential is measured by the relative amount of disruptive characteristics fulfilled. The result gives an indication of how threatening the potentially disruptive technology is likely to be at a given time.

In order to grasp this concept fully, one must be clear that the disruption phenomenon is a long-term process which can be divided into different phases (Christensen et al., 2004; Hüsig et al., 2005; Rafii & Kampas, 2002). As long as the disruption process is not completed, the monitored technologies might have different degrees of disruptive potential but the final outcome remains unclear until the end. Therefore, it might be misleading to measure disruptiveness of innovations, since disruption is an output condition or variable and not an input set of attributes per se. It might be of academic interest to find out how complete the victory of the entrants with the disruptive technology is, but that has few implications for an ex ante analysis when the disruption process is still in progress. For enhanced understanding

Figure 2. The phases of disruption – here in phase 7; The other six phases have already been successfully undergone by the entrant



the phases of disruption are described below and shown in Figure 2.

In Phase 1 the invention of the new technology with disruptive potential must be completed. Usually, technologies with disruptive potential are straightforward and the invention is at the architectural level (Christensen, 1992; 1997). Without an invention, no innovation whatsoever can follow. This frequently happens in the labs of the incumbents, who are characteristically unable to commercialize a potentially disruptive technology (Christensen, 1997). The first successful innovation phase is often carried out by some new firm that identifies a new untapped market need, or application by trail and error, which leads to Phase 2. The entrant's successful entry into an emerging market represents Phase 3. If the entrant can establish itself in this emerging market niche, it gains a beachhead to enter the mainstream market of the incumbent. If barriers of entry can be conquered effectively, the fourth stage of the disruption process has been reached. Now the entrant attracts customers away from the incumbent with typically lower unit prices and good enough performance features, which requires overcoming switching costs of the mainstream customers. In Phase 7 the incumbent starts to feel the loss and has the option of retaliation. If the counterattack remains unsuccessful due to the wrong set of resources, processes, and/or values in the changed context and is unable to develop or acquire them, the victory of the entrant with the alternative technology is complete (Christensen & Raynor, 2003; Overdorf & Christensen, 2000). With the total displacement of the incumbent in the market leader position by the entrant, the disruption is finalized in Phase 8. After this phase the disruptiveness of the process might be measured to determine how absolutely and completely the incumbents and their technology have been substituted.

Two other aspects should be considered when the concept of disruptive potential is used as a measure for ex ante analysis of the potential of emerging technologies to reach the end of the disruption process successfully. Due to the multiple uncertainties during this long-term development, the concept of disruptive potential must be seen as a heuristic concept that is based on historical analogies and pattern matching rather than primary number-driven quantitative analysis (Hüsig et al., 2007). Additionally, the disruptive potential of a technology can change over time depending on complementary innovations like business models. Therefore, the disruptive potential does not have a value that remains fixed over time, but needs to be reevaluated and permanently updated. These activities need to be embedded in a systematic innovation management process. In the next section it is elaborated on what such a management process and method for analyzing disruptive potential could look like.

MANAGEMENT PROCESS AND METHODOLOGY TO ANALYZE THE DISRUPTIVE POTENTIAL

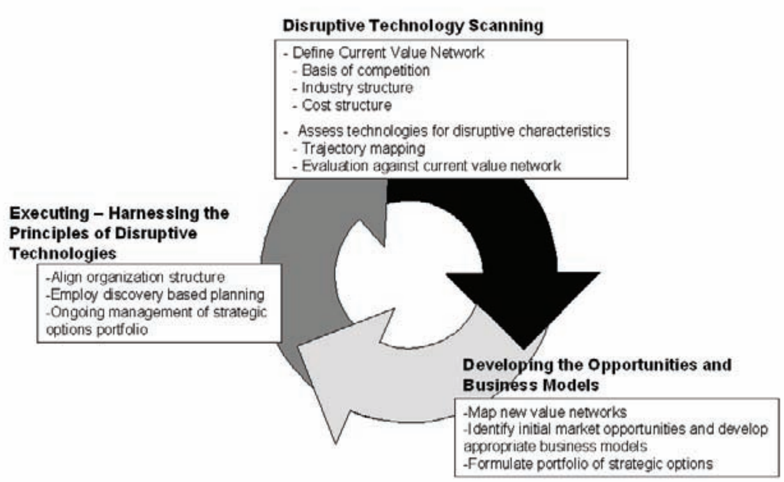
It is important to be able to distinguish between sustaining or disruptive technologies because of the different management actions which have to be taken in both cases. Unfortunately, as Christensen points out throughout his publications, it is not always easy to apply the categories of disruptive and sustaining innovation in practice: "Even people who deeply understood the theories [of disruptive innovation] struggled to use them in a repeatable and methodical fashion" (Christensen et al., 2004). Some of these difficulties might be due to the lack of terminological clarity, or methodological and formal deficits. One other reason for these difficulties is the fact that "disruption is a relative term" (Christensen & Raynor, 2003). This means that even though a particular innovation is disruptive to one player in an industry, it might be sustaining to another. This implies that firms have to be careful when categorizing innovations, particularly if most companies and public opinion consider an innovation to be of a disruptive nature. It highly depends on the individual strategy and business model of each firm, their resources, processes, and values.

Moreover, from an ex ante perspective there is the problem of technological and market uncertainty (Hüsig et al., 2005). To forecast the likelihood of disruption it is necessary to estimate the rate of technological progress for the performance supplied by the potentially disruptive technology and the technology in danger of disruption (Christensen, 1997). The market uncertainty inherent in the emergence of new sub-markets makes forecasts of potential customers, market growth, and size, or even useful applications of the potentially disruptive technology, very difficult (Chesbrough, 2003; Christensen, 1997).

However, even with all of these limitations in mind the concept of disruptive potential can add more orientation in this uncertain and complex undertaking which needs to be improved, comparing it with the former track record of innovation and technology forecasting in the telecommunications industry (Chandler, 1999; Christensen et al., 2004; Hüsig et al., 2005; Schoder, 2000). Given the high market and technology uncertainty in disruptive circumstances, most approaches provide mostly qualitative and heuristic methods like those presented by Christensen et al. (2004), Hüsig et al. (2005), or Rafii and Kampas (2002). Typically, the favored methodical basis for the approaches given in the literature so far consists mainly of checklists with indicators, scoring models, and trajectory mapping. As long as there is no comprehensive and easily applicable methodology to identify disruptive technologies from an ex ante perspective, the danger of misinterpretation and oversimplification will persist in practice. Therefore, in the following one methodology for analyzing the disruptive potential based on Christensen's research and the complements suggested by Adner (2002) to improve the precision of the disruptive technology concept and its usability for identification and forecasting developed by Hüsig et al. (2005) is presented together with a few improvements here. This approach helps managers, policy-makers and analysts to understand Christensen's theory better and to enable them to apply the insight more easily in the difficult task of identifying potentially disruptive technologies.

The method for estimating the disruptive potential of emerging technologies developed by Hüsig et al. (2005) must be embedded in a broader analysis and management process for disruptive innovations, which is shown in Figure 3. This

Figure 3. Analysis and management process for disruptive innovations (Hüsig et al., 2005)



method was developed to support the first phase of this process: Disruptive technology scanning. After the disruptive potential is estimated in the first phase, more specific managerial actions can be taken in the second and third phases if the disruptive potential has been shown to be high. In this article we will concentrate on the first phase, in which the disruptive potential is primarily analyzed.

The first step in this methodology is to define the current value network. The concept of the value network is described as the context within which a firm identifies and responds to customer needs, solves problems, procures input, reacts to competitors, and strives for profit (Christensen & Rosenbloom, 1995). Within such a value network, each firm's competitive strategy, and particularly its past choices of markets, determine its perceptions of the economic value of a new technology. These perceptions shape the rewards different firms expect to reap through the pursuit of sustaining and disruptive technologies. In established firms, expected rewards, in their turn, drive the allocation of resources towards sustaining technologies and away from disruptive ones. Inside a value network, the way value is measured differs across the network. Through their unique rank ordering of the importance of various product or service performance attributes, the metrics of value define in part the boundaries of the value network. These metrics of value can be used to develop the dimensions for the trajectory mapping in the manner of Figure 1. From a practitioner's perspective, the next question to be dealt with is the timeframe of the trajectory mapping. Since all factors influencing the firm are subject to certain conditions at a certain time, a planning and forecasting horizon has to be set for an appropriate analysis of the trajectories of performance improvement. In the case of disruption processes a longer time frame (at least the life-span of the potential disruptive technology) needs to be considered.

Subsequently, cost structures and business models of the incumbents must be analyzed and understood. Typically, the business models of the incumbents in a traditional industry like telecommunications do not differ too much from each other; therefore an "average" incumbent could be taken. For the individual analysis of the impact of a potential disruptive technology on a specific incumbent the subject of analysis is clear – and a deeper analysis of resources, values and processes is more likely to be feasible. Then the technology with potentially disruptive characteristics is evaluated in connection with the current value network. Depending on the stage of the disruption process, further elements of the emerging value network of the entrant can also be analyzed. To support this evaluation a number of consistent characteristics were extracted that typically indicate the threat of disruption and presented in a questionnaire. This questionnaire was further developed into a checklist containing the main indicators to analyze the disruptive potential (see Table 2). Often selecting an interdisciplinary sample of experts is suggested to ensure that both the market and technological views are covered (Hüsig et al., 2005). Others suggest to combine the analysis of potential disruption with other forecasting methods like using a large sample of experts via a Delphi study or use Business Model Ontologys (BMO) to compare the different player's business models (Osterwalder et al., 2005). It might also be helpful to complement the interview data with workshops, group discussions and additional studies to enhance the information basis of the analysis. Then all available information must be used to make decisions on the disruptive characteristics if possible. This task, the interviews and the moderating of the workshops should be performed by an expert in disruptive technology theory.

In the approach of Hüsig et al. (2005) the disruptive potential is measured by the relative number of disruptive characteristics fulfilled. Since this is simply a qualitative heuristic approach, more accuracy would seem rather misleading at this point. The results of the characteristics analyzed (yes = disruptive characteristics can be identified, no = disruptive characteristics cannot be identified, unknown = disruptive characteristics cannot be evaluated) should provide an estimate of how threatening the potentially disruptive technology is likely to be. In comparison with the method of Rafii & Kampas (2002) in which the factors can be rated on a seven-point scale that rates each contributing factor's disruptiveness, here no average positions can be taken or must be declared as unknown. The relative numbers of fulfilled disruptive characteristics equal the degree of the disruptive potential of the technology analyzed. The relative numbers of unknown disruptive characteristics equal the degree of uncertainty on the disruptive potential of the technology analyzed. The more advanced the disruption process is, the lower the degree of uncertainty concerning the disruptive potential of the analyzed technology might be, and vice versa.

This forecasting method is intended as a guideline for managers who are already familiar with Christensen's theory, and to support guided interviews with experts to gather and analyze the necessary facts in a structured way. The value added of this method is to help practitioners render the process of analyzing disruptive potential more systematic and comprehensive. Therefore, a more structural approach should also improve the quality of forecasting outcomes, especially if the number of the interdisciplinary experts using the proposed method grows and an aggregated analysis can be created. However, this methodology could also be combined with other frameworks and methods like Osterwalder et al. (2005) suggest. Especially in case of the analysis of the value network additional frameworks and tools might be needed to get the required input for the checklist anyway. Additionally, the proposed method should be developed further to expand the potential number of users.

POTENTIAL DISRUPTIVE TECHNOLOGIES IN TELECOMMUNICATIONS ANALYZED EX ANTE

As already mentioned above, there have been a few historic examples of successful disruptions in the telecommunications industry, especially in the equipment area, such as Cisco's development of the router or ROLM's creation of the PBX (Private Branch eXchange) market (Christensen et al., 2004). However, none of these innovations led to a widespread change in the telecommunications services. Therefore, much emphasis in recent research has been placed on the perspective of incumbent service providers (fixed and mobile) like Verizon, AT&T, Deutsche Telekom, BT, Vodafone, China Mobile, NTT, etc. who could be in danger of being disrupted by new entrants with technologies like VoIP, WLAN (Wireless Local Area Network or IEEE 802.11), or cable telephony, etc. (Christensen et al., 2004; Hüsig et al., 2005; Mannings & Cosier, 2001; Osterwalder et al., 2005; Rao et al., 2006). Historically speaking, the incumbents in telecommunications were highly vertically integrated, owned large established networks, and enjoyed high gross-margins while providing a high quality, mainly voice-based service (Christensen et al., 2004). Therefore, recent changes coming from the data and Internet industry as well as disintegration tendencies have been drawing the attention of industry, policy makers and researchers as a prime source of potential disruptions for the telecommunications service providers. The paradigm shift from centralized-wired-voice to decentralized-wireless-data could render traditional incumbents increasingly irrelevant. On the other hand, highly interdependent networked industries like telecommunications with significant network effects, in which regulation, and scarce resources play an important role, typically provide better avenues for co-option in favor of the incumbent telecommunications firms. Therefore, a more specific analysis for each potential disruptive technology is required. In this section a few published examples of potential disruptive technologies in telecommunications which were analyzed ex ante are presented, and a deeper look at the WLAN technology is taken.

Examples of Potential Disruptive Technologies in Telecommunications

There are already some examples of potential disruptive technologies in the telecommunications industry which have been analyzed ex ante in the academic literature. Some of these examples are given in Table 1. All of these examples are analyzed from the perspective of the incumbent service providers. In the cases of cable telephony, VoIP and wireless voice, the most threatening incumbents are the fixed line incumbent service providers. Wireless data/WLAN is typically seen as a threat to mobile service operators (MNOs). The results of the various authors were summarized when the studies were published and compared to the status quo in 2008. So far the track record of the ex ante studies looks quite good from today's perspective, although – even though a couple of years have passed – some of the developments are still unclear, such as the effects of instant messaging or VoIP on the telecommunications incumbents. This underlines the long-term process characteristics of the disruption phenomenon. However, most studies do not pay much attention to the geographic dimension and assume that the disruptive potential is similar everywhere, or concentrate implicitly on the USA market. This might underestimate different institutional conditions which influence the disruption potential (Chesbrough, 1999, 2003; Hüsig et al., 2007). In the next section the case of the WLAN technology is discussed to illustrate an example of a potential disruptive technology in telecommunications which was analyzed ex ante in greater depth.

The Disruptive Potential of WLAN for the MNOs

Various authors from business and academia have suggested that there is a disruptive potential of WLAN technologies for MNOs (Hüsig et al. 2005; Christensen et al. 2004). In particular, WLAN could be disruptive for incumbent MNOs' third generation (3G) data services, since they enable wireless Internet service providers (WISPs) to offer short distance, high bandwidth data services in heavily trafficked areas known as "hotspots". This new entrant WISPs could partly bypass the MNOs' data services based on licensed band mobile communication standards like GSM, GPRS or UMTS. These studies have frequently taken an ex ante perspective, since the public hotspot

Table 1. Examples of academic studies of potential disruptive technologies in telecommunications

Analyzed Technology/Innovation	Analyzed Disruptive Potential	Analysis by	Disruptive Potential To Date (2008)
Wireless Voice	High but co-opted by incumbents	Christensen et al. (2004)	High but co-opted by incumbents
Voice over Internet Protocol (VoIP)	High but co-opted by incumbents	Christensen et al. (2004) Osterwalder et al. (2005) Rao et al. (2006)	Unclear
Cable Telephony	Low, more sustaining	Christensen et al. (2004)	Low, more sustaining
Wireless Data/WLAN	Average, co-option by incumbents most likely	Christensen et al. (2004) Hüsig et al. (2005)	Co-opted and dominated by incumbents in most markets
Instant Messaging	High	Christensen et al. (2004)	Unclear

market has been developing rapidly. The study by Hüsig et al. (2005) used the method elaborated in this chapter. The authors came to the conclusion that WLAN was not very likely to represent a disruptive technology for the incumbent MNOs. This conclusion was drawn by a rigors analyses process which included the following steps:

1. Information and data gathering process using the checklist as guideline for interviews of interdisciplinary experts and to gather additional required data from market reports and various studies.
2. The value network and the metrics of value in mobile communications at the time of analysis were specified.
3. The basic characteristics of the WLAN technology and the mobile communication standards and their evolution over time were described.
4. The performance demand characteristics for wireless data services and applications and their evolution over time were qualified and quantified.
5. A trajectory map with both technologies together with the performance demand characteristics in the time-span of 1997-2005 was developed.
6. The checklist was used to analyze the gathered data and develop the final result for the disruptive potential.

The result of this process is summed up here briefly. The analysis of the disruptive potential using the methodology described above provided an overall disruptive potential score of 10 disruptive vs. 21 non-disruptive criteria and 7 unknown criteria. The key rationale for this conclusion is presented as follows. The conclusion that the likelihood of WLAN disruption for the MNOs was low was mainly due to the superior performance attributes in bandwidth, one of the most valued attributes for data services and the targeted customer group – business customers, who represent the most profitable segment in the value network of the threatened MNOs. The viewpoint was taken by Hüsig et al. (2005) that the existing WLAN based services at that point of time were not aimed particularly at downmarket segments with modest performance requirements but at the upmarket segments of business customers with high performance requirements, which does not indicate a typically disruptive pattern. From this viewpoint, WLAN enabled the existing consumers on the mainstream market to do things they could do previously, but not as well and probably less conveniently given the limitations of existing MNO services like WAP with low data rates when they wanted to connect to the Internet outside the office. Therefore, there were only underserved but no overserved customers in the mainstream market of MNOs, providing a potential for a radical sustaining upmarket innovation (Hüsig et al., 2005). Since almost everyone is an MNO customer, potential hotspot users and MNO customers have a high overlap, which leads to the conclusion that WLAN technology did not create a new customer group. Moreover, as accessing mobile data with high speeds has been valued by the high end customers, but not sufficiently enabled by existing MNO technologies and services, WLAN did not create an entirely new or low-end application for mobile data communication. Following this line of argument and taking the theory of disruption into account, Hüsig et al. (2005) concluded that in the case of WLAN no different levels of customer pressure and therefore no asymmetrical motivation between incumbent and entrant should exist, which favors a sustaining scenario for the WLAN technology. It was predicted that the appeal of WLAN for the incumbent MNOs' most valuable customers would force the MNOs to invest in this radical sustaining technology rather than to ignore it. Furthermore, without asymmetrical motivation, entrants' prospects seem to be rather limited in the long run given the wealth of the MNOs' resources which could be brought into the competitive arena. So far only limited empirical evidence from the hotspot market is available, but in most countries primarily the incumbent MNOs have taken advantage of the opportunity provided by the WLAN technologies and the hotspot market (Hüsig et al., 2007). This would support the results from the ex ante analysis performed with the methodology presented above.

FUTURE TRENDS AND CONCLUSION

In this article a method for assessing technologies for their disruptive potential from an ex ante perspective in the telecommunication industry was presented based on research by Hüsig et al. (2005) using the fundamentals of the disruptive technology concept pioneered by Christensen (1997), and recent theoretical developments. Even with the support of this methodology, it remains difficult to analyze the telecommunications sector using this theoretical underpinning as network effects, regulation, and scarce resources (e.g. frequencies) often play an important role in the adoption and diffusion of new services and technologies (Hüsig et al., 2005; Mannings & Cosier, 2001; Schoder, 2000). However, recent research in the telecommunications sector using the disruptive innovation framework shows the feasibility of the methodology in this industry (Christensen et al., 2004; Hüsig et al., 2007; Hüsig et al., 2005; Osterwalder et al., 2005; Rao et al., 2006). Major differences seem to be rooted in the fact that the incumbents in highly interdependent networked industries provide better avenues for co-option in favor of the incumbent telecommunications firms (Christensen et al., 2004). Moreover, this method should be seen as part of a broader analytical process, in which it supports the first disruptive technology scanning phase. The degree of disruptive potential can be measured by the relative number of disruptive characteristics fulfilled. The result gives an indication of how threatening the potentially disruptive technology is likely to be at a certain point in time, especially if a large group of interdisciplinary experts can be surveyed in order to provide an aggregated analysis (Hüsig et al., 2005). Finally, this method is especially useful as a guide for users who are already familiar with Christensen's theory and who wish to conduct interviews with experts from different domains to gather and analyze the necessary facts in a structured way. The value added of this method is to help managers, policy-makers and analysts systematically forecast the progress of disruptive technologies. An interdisciplinary group of experts is suggested to ensure that both the market and technological aspects are covered. In the case of telecommunications also the institutional and legal aspects might require special attention, so that also this type of experts must be considered.

There are some limitations that may hinder the reliability of this methodology, such as the quality of the input information provided, or the different emphasis with which certain aspects of the disruptive attributes are taken into account. The choice of performance criteria can also influence the result, as in most value networks the highest rank-ordering criteria are difficult to identify and technological trade-offs between performance criteria exist. However, as the studies on disruptive potential in telecommunications presented here indicate, most of the limitations can be successfully managed and useful ex ante analysis results can be provided.

Future research in this area should develop these methods further and improve the theoretical underpinning of the disruptive technology concept. From an empirical viewpoint, existing studies on potential disruptions in telecommunications need to be updated and monitored to evaluate how effectively they help managers, policy-makers and analysts systematically forecast the progress of disruptive technologies.

Table 2. Checklist with the main indicators for analyzing the disruptive potential modified from Hüsig et al. (2005)

Disruptive Potential Checklist	Yes	No	Unknown
Is there a technology with disruption potential?			
Does your firm have a potentially threatened technology (PTT)?			
Is the PDT (potentially disruptive technology) based on an architectural innovation?			
Is the PDT technologically straightforward, consisting of standard components?			
Does the PDT underperform with regard to most-valued performance dimensions?			
Are products based on a PDT already applied outside the mainstream market and your firm?			
Have prototypes or products based on the PDT been developed in your firm?			
Is the PDT aimed more at unpredictable than predictable needs of existing customers?			
Are the PDT initial cost/benefit characteristics not competitive with existing customers?			
Have existing customers shown a negative response to products based on the PDT?			
Are the products based on the PDT only useful for mainstream customers in a limited way?			
Are the customer needs or requirements regarding the PDT largely ill-defined or unknown?			
Does the PDT also attract less skilled consumers?			
Does the PDT also attract less wealthy consumers?			
Do the incumbents doubt whether there actually is a market for the PDT?			
Are smaller margins expected for the products based on the PDT?			
Are the products based on the PDT simpler, cheaper, more reliable or more convenient?			
Do the products based on the PDT have a worse cost- or price-performance ratio?			
Are the performance attributes of the products based on the PDT a disadvantage?			
Have the projects for putting the PDT on the market been stopped?			
Are there new firms that offer products based on the PDT?			
Have these new firms successfully created new markets for the PDT?			
Can these new firms achieve an attractive rate of return with products based on the PDT?			
Do the performance attributes based on the PDT develop faster than the new demand?			
Do the performance attributes based on the PDT develop faster than the old demand?			
Are the first products based on the PDT aimed primarily at down-market segments?			
Do the established firms reject introducing products based on the PDT?			
Has your firm reacted by introducing products based on the PDT onto the existing market?			
Are these products based on the PDT primarily aimed at customers in the existing market?			
Are your products based on the PDT marketed separately from an independent organisation?			
Are there no principal reasons that could impede the implementation of the PDT?			
Are there changes in the most-valued performance attributes on the mainstream market?			
Is there less willingness to pay for an increase in mainstream performance attributes?			
Are there customer groups with different demands regarding performance attributes?			
Does the PTT overshoot the performance requirements of the less-demanding customers?			
Does the PDT overshoot the performance requirements of the less-demanding customers?			
Does the PDT's trajectory of performance supply intersect with the trajectory of performance demand at the low end of the established market?			
Does the PTT undershoot the performance requirements of the new customers?			
Overall Disruptive Potential Score			

REFERENCES

Adner, R. (2002). When are Technologies Disruptive? A Demand-Based View of the Emergence of Competition. *Strategic Management Journal, 23*, 667-688.

Adner, R., & Levinthal, D. (2001). Demand Heterogeneity and Technology Evolution:Implications for Product and Process Innovation. *Management Science, 47*(5), 611 –628.

Bores, C., Saurina, C., & Torres, R. (2003). Technological convergence: A strategic perspective. *Technovation, 23*(1), 1-13.

Bower, J. (1970). *Managing the Resource Allocation Process.* Homewood, IL.: Irwin.

Burgelman, R. (1983a). A model of interaction of strategic behavior, corporate context, and the concept of strategy. *Academy of Management Review, 3*(1), 61-69.

Burgelman, R. (1983b). A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly*(28), 223-244.

Chandler, G. (1999). How to set up a forecasting system in telecommunications industry. *The*

Journal of Business Forecasting Methods & Systems, 18(2), 3-6.

Chesbrough, H. W. (1999). The organizational impact of technological change: a comparative theory of national institutional factors. *Industrial and Corporate Change, 8*, 447-485.

Chesbrough, H. W. (2003). Environmental influences upon firm entry into new sub-markets: Evidence from the worldwide hard disk drive industry conditionally. *Research Policy, 32*(4), 659-678.

Christensen, C. M. (1992). Exploring the Limits of the Technology S-Curve. *Production and Operations Management, 1*(4), 334-366.

Christensen, C. M. (1993). The Rigid Disk Drive Industry: A History of Commercial and Technological Turbulence. *Business History Review, 67*(4), 531-588.

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston/Massachusetts: Harvard Business School Press.

Christensen, C. M., Anthony, S. D., & Roth, E. A. (2004). *Seeing What's Next - Using the Theories of Innovation to Predict Industry Change*. Boston: HBS Press.

Christensen, C. M., & Bower, J. (1996). Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strategic Management Journal, 17*, 197-218.

Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution: Creating and sustaining successful growth*. Boston, Massachusetts: HBS Press.

Christensen, C. M., & Rosenbloom, R. S. (1995). Explaining the Attacker's Advantage: The Technological Paradigms, Organizational Dynamics, and the Value Network. *Research Policy, 24*, 233-257.

Danneels, E. (2004). Disruptive technology Reconsidered: A Critique and Research Agenda. *Journal of Product Innovation Management, 21*(4), 246-258.

Gilbert, C. G. (2002). Beyond Resource Allocation: Towards a Process Model of Response to Disruptive Change. *Harvard Working Paper #03-018*.

Gilbert, C. G. (2005). Unbundling the Structure of Inertia: Resource versus Routine Rigidity. *Academy of Management Journal, 48*(5), 24.

Gilbert, C. G., & Bower, J. (2002). Disruptive Change - When Trying Harder Is Part of the Problem. *Harvard Business Review,* (5), 94-101.

Govindarajan, V., & Kopalle, P. K. (2006a). Disruptiveness of innovations: measurement and an assessment of reliability and validity. *Strategic Management Journal, 27,* 11.

Govindarajan, V., & Kopalle, P. K. (2006b). The Usefulness of Measuring Disruptiveness of Innovations Ex Post in Making Ex Ante Predictions. *Journal of Product Innovation Management, 23,* 7.

Hüsig, S., Dowling, M., & Hipp, C. (2007, 1-3 June). *Analysing the Disruptive Potential of WLAN: A Three-Country Comparison of the Hotspot Market*. Paper presented at the 5th International Symposium on Management of Technology, Hangzhou, P.R. China.

Hüsig, S., Hipp, C., & Dowling, M. (2005). Analysing Disruptive Potential: The Case of W-LAN and the Mobile Telecommunications Network Companies. *R&D Management, 35*(1), 17-35.

Lia, F., & Whalleyb, J. (2002). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy, 26*(2), 451–472.

Lyytinen, K., & King, J. L. (2002). Around the cradle of the wireless revolution: the emergence and evolution of cellular telephony. *Telecommunications Policy, 26*(3-4), 97-100.

Mannings, R., & Cosier, G. (2001). Wireless everything — unwiring the world. *BT Technol J Vol 19 No 4 October 2001, 19*(4), 65-76.

Osterwalder, A., Ondrus, J., & Pigneur, Y. (2005). *Skype's Disruptive Potential in the Telecom Market: A Systematic Comparison of Business Models.* Retrieved 07.04., 2008, from http://www.hec.unil.ch/yp/Pub/05-skype.pdf

Overdorf, M., & Christensen, C. M. (2000). Meeting the Challenge of Disruptive Technology. *Harvard Business Review*, 67-76.

Pfeffer, J., & Salancik. (1978). *The External Control of Organisations: A Resource Dependance Perspective*. New York: Harper & Row.

Rafii, F., & Kampas, P. J. (2002). How to identify your enemies before they destroy you. *Harvard Business Review, 80*(11), 115–123.

Rao, B., Angelov, B., & Nov, O. (2006). Fusion of Disruptive Technologies: Lessons from the Skype Case. *European Management Journal, 24*(2-3), 174-188.

Schoder, D. (2000). Forecasting the success of telecommunication services in the presence of network effects. *Information Economics and Policy, 12*(2), 181-200.

KEY TERMS

Disruptive Potential: Concept and measure for ex ante analysis of the potential of emerging technologies to reach the end of the disruption process successfully. The degree of disruptive potential is measured by the relative number of disruptive characteristics fulfilled. The result gives an indication of how threatening the potentially disruptive technology is likely to be at a certain point in time.

Disruptive Technologies/Innovations: Technologies or innovations which disrupt an established trajectory of performance improvement downwards, or redefine what performance means. In the long run the initially inferior technologies from new markets or low end market segments typically introduced by new entrants become good enough to meet the demands of the mainstream customers of the incumbents, so that the incumbents are forced toward higher margin customer groups or out of the market entirely.

Entrants: New firms in a certain value network. Often entrants introduce potentially disruptive technologies in traditional value networks. Examples in telecommunications are Skype or Cisco.

Incumbents: Traditional and typically large firms which operate in an existing value network. Frequently their successfully manage sustaining innovations. Examples in telecommunications are BT, Vodafone, NTT, France Telecom or Deutsche Telekom.

Innovation: Changes in the technology of an organization.

Sustaining Technologies/Innovations: Technologies or innovations which have a sustaining impact on an established trajectory of performance improvement and give the known customers more and better versions of what they have valued up to now. They lead to innovations that allow the firms to provide their existing customers with better and more profitable products. Incumbent firms almost always ultimately master this type of innovation but are often unable to cope with disruptive ones.

Technology: The process by which an organization transforms labor, capital, materials, and information into products and services.

Value Network: The context within which a firm identifies and responds to customer needs, solves problems, procures input, reacts to competitors, and strives for profit.

Chapter VIII
Convergence of the Internet and Telecommunications

John B. Meisel
Southern Illinois University Edwardsville, USA

Timothy S. Sullivan
Southern Illinois University Edwardsville, USA

ABSTRACT

The local exchange telephone market is no longer considered to be a natural monopoly as a result of technological advancements. In particular, the power, versatility, and adaptability of digital technology has enabled traditional telecommunications companies to transition their network architecture from one based on copper wires and circuit switches dedicated to the provision of high quality transmission of point-to-point voice signals to a multi-faceted, general network based on fiber optic cables and packet switches capable of providing an array of voice, data, and video services. A network layers model is introduced to conceptualize the process of convergence to the Internet model. Convergence is fundamentally changing the nature of what it means to be a telecommunications company and promises to alter the market structure of many voice, data, and video markets. Convergence has pitted wireline telecommunications companies in a fierce rivalry with cable companies for voice, data, and, increasingly, video services. More generally, this research provides a framework to explain the convergence of communications networks and identifies and analyzes key issues that confront public policymakers. One key competition issue, termed network neutrality, addresses the concern that the evolving broadband network architecture will enable network providers to favor the provider's services or affiliated services at the expense of independent rivals.

INTRODUCTION

At the close of the 20th century, a traditional, prototypical telecommunications company provided high quality point-to-point voice services utilizing a technology specially designed for transmitting voice signals. This wireline architecture was comprised predominantly of copper-based local loops, circuit switches, and local transport and long distance conduits that incorporated some fiber-optic technology. In such centralized phone networks, intelligence resided in the core of the network in the switch technology, not in the end devices connected to the network as it does in the decentralized Internet architecture. In such a network, the provision of the service, voice, was closely identified with the underlying physical infrastructure that provided the service. For much of the 20th century, United States telecommunications companies were thought to be natural monopolies in the provision of voice services and thus subject to regulation in dual jurisdictions, at the state and federal levels. More generally, regulation of electronic communications networks was based on the premise that a particular service (such as voice, data, or video) was inextricably linked to a specific underlying infrastructure used to provide the service.

With technological improvements in digital and network technology and the spectacular growth of broadband technology and the Internet at the end of the 20th century, the traditional architecture of phone networks (and other communications networks) and the scope of regulation of telecommunications companies are gradually changing. Telephone companies are transforming their provision of voice and other broadband services from copper-based networks to advanced fiber-optic networks. The purpose and scope of this research is to describe and analyze the restructuring of telecommunications networks and telecommunications companies as they respond and adapt to technological and economic changes.

As telecommunications/entertainment/information networks converge to an advanced, multifaceted architectural broadband platform that is capable of offering an integrated bundle of voice, data, and video broadband services, it is incumbent on regulatory authorities to change the nature of the regulatory approach toward these telecommunications companies. A pattern of deregulation is in process but there still exist unanswered questions regarding the appropriate regulatory response to convergence. Convergence is creating new competition/regulatory issues in a variety of formerly distinct industries. Peha, Lehr, and Wilkie (2007), three leading Internet scholars, argue that "At root, the debate is about the future of regulatory policy for the communications sector as we move toward a post-convergence world where legacy industry boundaries and regulation of cable television, broadcasting, and telecommunications need to be reconciled with the challenges of competition among broadband platform services, based on the Internet" (p. 710). For instance, as the telecommunications network converges on the Internet model, telephone companies have discussed implementation of a two-tiered (non-neutral) Internet that would replace the standard practice by Internet service providers of treating generally all kinds of packets in the same manner (i.e., neutrally) regardless of the application or provider. Observing a neutrality principle implies that broadband access providers do not favor some packets at the expense of others, that is, there exists a lack of discriminatory behavior toward applications and content utilizing the network.

A two-tiered Internet can be viewed as a strategic maneuver by vertically integrated broadband service providers to gain a competitive advantage in the market for Internet services and applications. Richard Notebaert, Chief Executive Officer of Qwest Communications, for example, argues that telephone companies should be able to negotiate commercial agreements with content providers such as Google or Amazon to give Internet companies an advantage over rival firms because, "we're all trying to provide a little bit of differentiation for a competitive edge. That's what business is about" (Reardon, 2006a). With a non-neutral Internet, broadband access providers install intelligence within the network in order to favor some packets and disfavor others and, hence, exert control over applications and content utilizing their network. The controversy over implementation of a non-neutral platform is captured in the term net neutrality. Potential competitive issues arising from the design and implementation of a two-tiered broadband architecture involve (1) a telecommunications company's treatment of an emerging Internet service (such as Voice over Internet Protocol (VoIP)) that competes directly with the company's core, legacy service, (2) a telecommunications company's treatment of its newly created or purchased/favored affiliated Internet content and/or Internet application service that competes with a non-affiliated company's substitute service, and (3), more generally, the ability and incentive of a telecommunications company to attempt to leverage market power in broadband access into Internet services dependent on a broadband platform. Issues involving the concept of net neutrality are examined using an antitrust framework. Such a policy framework is appropriate since technological and economic changes and the convergence of previously distinct network architectures (such as the telephone and cable networks) to the advanced broadband Internet model is increasing the number of competitors facing telephone companies. The era of telecommunications natural monopoly is over and the advent of inter-platform competition is here.

The next section explains the term convergence and compares and contrasts the current telephone network with the broadband Internet network and describes the evolution of the telephone network to the advanced Internet broadband architecture using the concept of the network layers model. The following section describes the changing nature of the external environment confronting telephone companies and explains how these companies are adjusting to technological convergence, regulatory changes, and economic competition. The remaining sections identify and analyze competition issues involving broadband service providers that result from the convergence to the Internet broadband architecture and provide a summary.

CONVERGENCE

For most of their existence, telecommunications companies were viewed as specialized providers of voice services using a unique copper-based technology. That is, a telephone company's network architecture was dedicated to provide analog voice services and the physical components of the network included circuit switches, copper local loops, and high capacity transport and long distances lines. The transition to a new broadband architecture based on the Internet protocol is underway. The transitional network architecture is characterized by local loops, part copper and part fiber, carrying a combination of digital and analog signals. Signals today reach their final destination by traveling through either circuit switches and/or packet switches in local and long distance transmission links increasingly built with fiber. This transitional architecture enables telecommunications companies to provide a combination of voice and data, especially narrowband and broadband Internet access, services. The current architecture is likely to evolve to an all fiber-based, digital packet-switched broadband network capable of transmitting voice, data, and video services. The transition and transformation of telecommunications networks can be described and analyzed using a network layers conceptual model. (See, for example, Nuechterlein & Weiser, 2005, Chapter 4; Whitt, 2004 and Werbach, 2002.)

Network Layers Model

The network layers model is an analytical tool that views a communications network as a set of discrete, functional layers. That is, the concept of a layered network means unbundling the complex task of provisioning a communications signal into smaller, discrete subunits, each of which is designed to address a specific task. Each layer in the network contributes to the provision of a communications service within a hierarchical framework. Each layer has interfaces that facilitate its coordination with adjacent layers. Each distinct layer of a communications network can be examined to assess the lack of or degree of horizontal competition from competing networks within a given layer (e.g., competing instant messaging applications, competing email applications, competing infrastructure providers) as well as its vertical relationship to adjacent layers. Table 1 provides an overall schematic of the model.

The lowest layer (layer 1) is called the physical access layer. (See, for example, Whitt, 2004, and Weiser, 2003.) This layer is associated with the physical infrastructure and includes elements such as personal computers, servers, smart devices, copper-based local loops, fiber-based local loops, circuit and packet switches/routers, digital subscriber lines (DSL), cable modems, coaxial cable, power lines, radio towers and antennas, and long distance trunks. This layer is responsible for the transmission of content between users within a network or across networks. Service providers at this layer provide two basic services: (1) local shared transport (transmissions from end office switches to tandem switches) and long haul backbone services which involve high capacity lines connecting tandem switches and network access points that interconnect different networks, and (2) local access services which involve the provision of first-mile and last-mile dedicated (or shared among a small number of customers) connections between end offices and local residences and businesses. Service providers at layer 1 include telephone companies, cable companies, and wireless companies.

On top of the physical access layer sits the logical layer (layer 2). This layer in essence identifies the common language that enables different computers around the world to communicate. In other words, this layer manages how information travels over the Internet. It involves a set of

Table 1 Schematic of the network layers model

Layers	Typical Providers
Content Layer Video, text, audio, voice	Entertainment companies, information companies such as NBC, ABC, CBS, ESPN,CNN, FOX
Applications Layer Email, peer-to-peer, instant messaging, electronic commerce, search, voice over Internet protocol, video hosting, video conferencing	Developers of software such as Google, eBay, Skype, Yahoo, YouTube, Vuze, BitTorrent, Microsoft
Logical Layer Open set of protocols, such as TCP/IP protocol suite	Broadband service providers, includes telephone and cable companies
Physical Layer Loops, switches, trunks, routers, cellular towers	Network providers such as Comcast, AT&T, Verizon, Qwest,Time Warner, Madison River Communications

intelligent protocols that lays out the ground rules for communication. Information is assembled into packets with a specific format that includes addressing rules, the Internet Protocol (IP), and rules to inform packet switches along the way, the transmission control protocol (TCP), where the packets want to go. In short, the Internet protocol packages and routes data. These protocols work on top of any configuration of the physical layer. Importantly, these protocols have become network software standards and lay in the public domain for use by any applications developers. Internet access providers provide access to the Internet and operate at layer 2.

The third layer in the model is the applications layer. Applications such as e-mail, the World Wide Web, search engines, instant messaging, and audio and video players have spurred the growth of the Internet. These applications reside as software on computers and servers. Applications are based on layer-two logical software for addressing and transmission and add their own innovative protocols for computers to exchange information. The creation of new applications provides incentives for users to use the Internet and to upgrade their broadband access to enjoy more bandwidth intensive and real-time applications like peer-to-peer file sharing and real-time video. Given the layered and modular nature of the Internet, applications developers need only concern themselves with writing code for new applications compatible with the IP protocol and not be concerned with also provisioning physical infrastructure and/or logical protocols. (Contrast this, for instance, with an inflexible centralized architecture such as the original telephone network, where, if a developer wished to provide a new application for the network, the application developer must seek permission from the network's owner to do so and must configure the application to the specific network's underlying physical infrastructure and logical code.) Application service providers include independent Internet companies such as Google, Amazon, and eBay as well as layer-1 and -2 vertically integrated service providers.

The fourth layer of the model is content. This is the actual information that is exchanged. Movies, music, voice communications, and data can all be digitized and exchanged between computers using the three lower layers. Content is what users are ultimately interested in exchanging with others. Commercial entertainment, information, communications companies, and increasingly today, amateur end users create content and thus operate at layer 4.

The wireline telephone network has been upgraded to provide voice and broadband services using a digital subscriber line (DSL) architecture. This means that telephone companies are simultaneously operating two networks, a dedicated circuit-switched voice network and more general packet-switched data network. This enables telephone companies to utilize their preexisting copper local loops to residences and businesses and to upgrade the services provided to these customers. Thus, a telephone company integrates layer 1 and 2 functionalities for it provides basic transmission and enhanced computer processing of data for the customer. A residential customer can use the local loop to both make voice calls that travel over the circuit-switched public telephone network and to access the Internet over a broadband connection using the packet-switched Internet to reach the intended destination. The likely goal of telephone companies is to continue the evolution of their networks to the Internet model in which they provide all services over the Internet platform, ultimately, including voice services.

Economic, Regulatory, and Technological Environments

Telephone companies are not alone in adapting to the inexorable trend toward technological convergence as their current network architecture transitions to incorporate the Internet layers model in order to offer additional services built on the underlying physical infrastructure. Operators of cable television systems are undergoing a similar transition. (It should be noted that operators of power lines, satellite systems, and private and municipal wireless systems are also creating or upgrading networks to provide enhanced broadband Internet access and upper layer services.) As well as providing multichannel video programming, cable companies today provide their customers broadband Internet access and voice communications, often using VoIP technology. That is, voice becomes an Internet application that rides on top of the logical and physical layers of a cable network. Cable companies are now in direct competition with telephone companies as "triple play" providers of broadband Internet access, voice communications, and multichannel video distribution. As of the end of 2006, there were five million subscribers in North America that purchased three communications services from the same company (which at this point is overwhelmingly a cable company). The predicted number increases to 10 million by 2010 (Britton & McGonegal, 2007). These competitors utilize different physical infrastructures but as they transition to the Internet layers model they are in a position to provide similar services that are built on the same logical IP protocol. This transformation of telephone and cable networks has forever severed the link between a specialized service and a specific underlying physical architecture.

As noted above, cable television and phone services are undergoing a remarkable convergence, which is likely to accelerate in response to technological changes. Recently, cable companies have made sizable inroads into the market for telephony service. At first, cable companies constructed physical facilities similar to the telephone companies' circuit-switched architecture but with technological improvements in broadband networks they have since found it more economical to provide voice services as an application utilizing the broadband networks they were constructing to compete in the provision of Internet broadband services. As of late 2007, cable companies provided telephone service over their broadband lines to over 12 million US households, or about 11% of the households that are capable of using cable voice service. Comcast, Cox, Time Warner and Cablevision all have more than one million voice subscribers, and virtually all major cable companies offer telephony services (Brumfield, 2007b). While the growth rate has slowed recently, all of these companies continue to add subscribers (Barris, 2007 and Brumfield, 2007a). When stand-alone services, such as Vonage and Skype, are included, over 16 million US households use the Internet for phone service, a number that has nearly quadrupled in less than two years (Mangalindan & Rhoads, 2007). While the stand-alone services have suffered both revenue and legal problems in recent years, cable companies have been fairly successful bundling phone service with other services, such as television and broadband (Shwiff, 2007).

While most telephone companies are unable to deliver standard broadcast signals over conventional phone lines, many are partnering with satellite television companies in anticipation of offering television service directly. Verizon, for example, has partnered with DirectTV to offer satellite television service to its residential customers. Similarly, AT&T has partnered with DISH Network, and the two are rumored to be discussing a buyout deal (Moritz, 2007). These partnerships are credited in part with recent declines in the number of video customers held by cable companies (Searcey, 2007).

We already are witnessing a rapid expansion of video content provided over the Internet. Currently all of the major US television networks and a sizable number of "cable" networks offer some portion of their primetime schedules online for free. Other programs are offered, for example, through Apple's iTunes for a fee. Other types of content, traditionally viewed on television, are being introduced at a rapid rate. The National Football League (NFL), for example, offers the SuperFan package, which allows DirectTV customers who purchase the Sunday Ticket package to watch NFL games live over the Internet. Major League Baseball's MLB.tv allows subscribers to watch live baseball games over the Internet. While relatively inexpensive hardware is already available that allows images on a home computer to be played on a standard television set (and *visa versa*), advances in LCD technology are likely to increasingly blur the line between television monitor and computer monitor.

Advances in broadband access and technology will mean that phone service and television service can both be provided by the Internet, allowing broadband service providers to replace both services regardless of its history as a cable or telecommunications company**.** Because of the sizable bandwidth requirements for video, adoption of the Internet as a household's video provider requires a broadband connection. As of late 2007, only 58 million US households are broadband subscribers, accounting for roughly 53% of US households (Malik, 2007). Further, many households currently opt for the slower (and less expensive) downstream speeds of less than 5.0 megabits per second (Mbps). However, as broadband access increases, especially access using far-faster fiber-optic cable, as opposed to standard phone lines and coaxial cable, it is easy to envision the Internet replacing traditional cable and satellite as the primary means of delivering video to US households. At this point, in addition to traditional Internet content, the household's broadband service provider will become the delivery mechanism for voice and video.

The race is on between these two converging industries to be the superior provider of broadband access service (and applications and content) and the early results are as follows:

- The top providers of broadband service are remarkably split between telecommunications and cable companies:
    - Cable, in total, has a slight overall edge (Grant, 2006).
    - The top providers are AT&T, with almost 14 million; Comcast, with almost 13 million; Verizon, with almost 8 million, and Time Warner, with over 7 million (Malik, 2007).
- In the short run, it may be more difficult for phone companies to enter the television market than for cable companies to enter the telephone market because phone companies must expand their networks for greater capacity, while cable companies can use their current infrastructure for some time. Experts are divided, however, whether the cable infrastructure will be able to support greater bandwidth without major investments in infrastructure.
- Verizon now offers fiber-optic service (FIOS) to Internet customers in 16 states. This fiber-optic network service is capable, in some states, of speeds of up to 50 megabits per second (Mbps). Other phone companies are pursuing a similar strategy, but are focusing on new housing developments (Grant, 2006).

As well as a trend toward technological convergence, telephone companies are active participants in creating a trend toward consolidation of telephone assets and a continuing pattern of telephone deregulation. AT&T was divested in 1984 into (1) a free standing long distance company (called AT&T) that would compete in a competitive market against companies such as Sprint and MCI, and (2) seven independent regional Bell operating companies (Pacific Telesis, U.S. West, Southwestern Bell, Ameritech, Bell South, Bell Atlantic, and NYNEX) that were viewed as natural monopolies in the provision of stand-alone local telephone service in their unique geographic areas and thus subject to continuing regulation. The economic theory in 1984 was that the long distance market was capable of competition safeguarding consumers while local telephone service was not, thus regulation was still necessary to protect consumer welfare and to prevent local companies from leveraging their local assets into long distance and enhanced services markets.

In the next quarter century and with the intervening passage of the Telecommunications Act of 1996, whose goal was, in part, to promote competition in local telephone markets, the industry has been restructured back toward the original AT&T model of a vertically integrated company. Today, there exist three remaining Bell companies, and AT&T and MCI have been merged into those companies. The demise of AT&T and MCI was mainly due to the rapid decline of the stand-alone long distance business model (in part, due to the entry of the Bell companies into long distance markets) and legal rulings that thwarted their attempts to enter local residential markets using unbundled elements of the conventional telephone network at regulated rates as competitive local exchange carriers. However, these two companies brought assets to their buyers that included long standing business relationships with high volume metropolitan corporate customers who demand sophisticated voice and data services.

The three remaining Bells, AT&T (the result of a 2005 merger between AT&T and SBC, which was itself a combination of three Bells, Southwestern Bell, Pacific Telesis and Ameritech, and Bell South), Qwest Communications International (which merged with U.S. West in 2000), and Verizon Communications (the result of mergers between Bell Atlantic and NYNEX and later MCI) are the major telecommunications providers of long distance service, local service and broadband Internet access. It is argued that the real, lasting competition to these wireline companies comes from four national network wireless companies (although two of the largest cellular providers are owned by Verizon and AT&T) and Voice over Internet Protocol (VoIP), an Internet entrant into the voice market (Picker, 2005).

Decisions by the Department of Justice and the Federal Communications Commission (FCC) to allow the mergers by AT&T and Verizon have been accompanied by certain regulatory conditions placed on the companies. Before examining several specific conditions from these consolidations, several of the most significant deregulatory efforts for telephone companies are in order. A major goal of the FCC since passage of the Telecommunications Act of 1996 has been to promote competition in the lowest layer in the network layers model, the physical infrastructure layer. The Act required the incumbent local exchange carriers to open up their local infrastructure in various ways (e.g., unbundle network elements, act as a wholesaler for their retail service) to facilitate the entry of competitive local exchange carriers. Inevitably, disputes between the incumbents and potential entrants over network access have accompanied FCC deregulation efforts. The Bell incumbents led the resistance to these efforts in court, eventually forcing the FCC to back off from its aggressive efforts to promote last-mile intramodal competition in the traditional telephone network. As of 2007, FCC regulatory efforts appear to be focused on providing incentives for incumbents to upgrade their physical networks, especially toward increased use of advanced technology such as fiber, and to encourage the entry of last-mile inter-modal competition in broadband markets.

At the same time, the FCC ruled that Internet access providers using cable modem technology and DSL technology to provide Internet access to residential customers should not be subject to common carrier regulation. DSL service had been subject to common carrier rules including the legal obligation for telephone companies to offer access to its physical transmission lines to the Internet to independent Internet service providers and to rate regulation of such access. The FCC could have decided to include cable companies offering Internet access under the same common carrier regulations. Instead, the FCC ruled that cable and telephone Internet access services should be classified as information services, not telecommunications services, and thus, free of Title II regulations in the Communications Act. The US Supreme Court, in 2005, in *National Cable & Telecommunications Association v. Brand X Internet Services*, upheld the FCC's classification scheme in its Brand X decision regarding cable modem service. Similarly, the Third Circuit upheld the FCC's decision that DSL service should be classified as an information service. Effectively, this means that telecommunications carriers and cable companies are not subject to common carrier obligations and, thus, can offer Internet access service, which under the network layers model bundles layers 1 and 2 (the physical layer and the logical layer), without the requirement to provide an open interface between these two layers to unaffiliated competitive Internet access providers. (The term "open access" refers to the ability of independent Internet access providers to lease from regulated tariffs the last-mile facilities (i.e., layer 1) of incumbent common carrier companies.) Thus, a pattern has been established that reduces, if not eliminates, access to the physical layer infrastructure as a method of attempting to move the market structure of the physical layer to a more competitive structure. Again, the FCC's rationale is based on the argument that regulation diminishes the incentive for incumbents and entrants to invest in advanced infrastructure technology and that regulatory efforts to spur intramodal competition by forced sharing of incumbent telephone companies' last-mile infrastructure is unfair to incumbents and, in the long term, futile in providing real competition.

The next step in the deregulation trend has eased regulations in the commercial broadband market. This involves companies that provide high-capacity broadband services to large corporate customers. The process for deregulation has been initiated by the telephone companies through use of a tool called a forbearance petition. This approach was formalized in the Telecommunications Act of 1996 and allows telephone companies to request relief from regulations that are no longer necessary due to the presence of competition. Verizon and AT&T have received regulatory relief in this market in response to such petitions (Boles, 2007 and Schatz, 2007). At this time, broadband Internet access to residences and small business, and large business are deregulated. One remaining market still subject to regulation and on the agenda for deregulation is the special access service market. Uses of special access lines include connecting large businesses in metropolitan regions directly to long distance carriers' points of presence and connecting cell towers to local exchange switches. AT&T is subject to regulation of the special access rate for four years as a condition of the AT&T-BellSouth merger. The issue of the extent of competitiveness of the special access market is still unresolved.

Conditions on the telecommunications mergers described above have played a role in the current controversy over the concept of network neutrality. The FCC wishes to maintain the "openness" of the Internet without having to write specific rules to control the behavior of broadband Internet service providers. At the end of 2006, about 95% of 82.5 million last-mile, high-speed lines (including 58.2 million residential lines) are owned by telecommunications companies and cable companies (Federal Communications Commission, 2007). Thus, there exists a concern that these companies may attempt to leverage their market power into competitive applications and content markets. In response to this concern, the FCC conditioned approval of the Verizon-MCI and AT&T-SBC mergers on promises from these companies that they would abide by the "openness" goal for a two year period and follow unwritten net neutrality rules. Specifically, the FCC issued a broadband policy statement that called for, among other things, preserving the neutrality of treatment of packets across the Internet by all broadband service providers (Federal Communications Commission, 2005).

More significantly, in the approval of the AT&T-BellSouth merger, a similar net neutrality condition was agreed to but with a telling loophole (Reardon, 2006b). AT&T (as well as Verizon) has been upgrading its network infrastructure in order to provide Internet Protocol Television (IPTV) to compete directly with cable and satellite for video viewers (Faulhaber, 2007). The loophole says that AT&T will not have to follow net neutrality principles on the IPTV network for it is a private/managed network but only follow them on their existing public broadband network. Again, this reflects a pattern by the FCC of trying to provide incentives for infrastructure companies to upgrade the technology of the network. Critics worry that the division of the Internet into two parts, a private/managed Internet and a public Internet will create competition problems in the future. Potential competition issues arising from a two-tier Internet is discussed in the next section of the chapter.

COMPETITION ISSUES

Telecommunications companies will generate revenue in broadband by (1) selling access to the public Internet (basic access service), (2) selling access to the private/managed Internet (premium access service) and (3) selling its own integrated or affiliated applications and content over the managed network. The antitrust concern is that incumbent broadband Internet access providers, to the extent that they possess market power in broadband access markets, will leverage that power in an anticompetitive manner into presumptively competitive applications and content markets under certain conditions.

In general, broadband access providers have a natural incentive to facilitate and encourage the efficient use of broadband platforms for complementary services. More complementary applications and content increase the value to consumers of the platform. The incentive by platform providers to encourage new complementary applications and content is called the principle of "internalizing complementary efficiencies" (Farrell and Weiser, 2003). In other words, broadband platform providers generally have no incentive to discriminate against unaffiliated providers of upper layer services for such a strategy is not in the broadband providers' best profit maximizing interest. However, exceptions to the principle exist, some of which may be relevant to telecommunications companies acting as broadband service providers and the plans they have, in a world of convergence, to integrate vertically into applications and content markets. Exceptions raise the potential for the threat of anticompetitive vertical behavior by broadband service providers toward unaffiliated applications and content providers.

In order to offer IPTV and other Internet services, telecommunications companies are in the process of making multi-million dollar fiber-based investments over a period of time to upgrade the supply capability of their broadband networks. Importantly, the way in which the next-generation network is designed (i.e., the architecture of the advanced broadband network) is likely to have long-lasting effects on the structure of Internet markets. According to Nuechterlein and Weiser (2005), "…once a monopolist in a network industry chooses a particular architecture -- say, one that facilitates certain applications for a limited set of preferred providers- network effects may cause the market as a whole to become dependent on that architecture, even if it turns out to have in severable anticompetitive characteristics" (p. 173). In short, network architecture acts as a causal factor for market structure. Network design can result from either technical considerations or from business motivations. As with any investment, the specific question becomes how telecommunications companies are going to generate sufficient revenue to pay off the investment.

One option is to offer, as an alternative to the best efforts public Internet, a better than best efforts private Internet. The private/managed Internet would have an enhanced capability to meet the needs of Internet services that require a higher quality of service than traditional Internet services such as email, web browsing, and small to medium-sized file transfers. This strategy is called vertical differentiation and attempts to provide a qualitative advantage to packet transmissions over the telephone companies' private Internet compared to the public Internet. Given the current (de)regulatory environment and after the two year moratorium on capitalizing on vertical differentiation as agreed to during negotiations to win approval of the recent mergers, telephone companies will have the commercial freedom and flexibility to charge applications and content providers for access to enhanced private IP delivery. Public statements by senior executives in the largest telephone companies have expressed a strong interest in capitalizing on this freedom. Moreover, technology (such as deep packet inspection) has advanced to the extent that the ability to discriminate exists through installing intelligence within the lower layers in the network to discover and identify the source, type, and destination of packets. This means that the broadband service provider can uncover information about the packets traveling on its physical infrastructure and dictate the quality of transport a packet is to receive based on this information.

The public Internet is coping with the increased heterogeneity of demand from new services. Streaming video and peer-to-peer video file transfers are bandwidth intensive services while VoIP and online gaming are sensitive to delay in delivery of packets and to variation in that delay. These new demands have strained the supply capability of the best effort, neutral treatment of all packets, public Internet. (Swanson, 2007, provocatively refers to the predicted tremendous increase in demand for broadband capacity as the coming "exaflood."). In order to address the needs of these new services, a discussion has ensued about the appropriateness of broadband access providers providing differentiated quality of service to applications and content providers, and end users.

In order to receive a return on their investment in the next-generation broadband network, telecommunications companies want to differentiate vertically the quality of Internet access they provide to applications and content providers to reach end users. Telecommunications companies plan on restricting access to the non-neutral Internet that provides a higher quality of service than the neutral Internet, and, then, offer a higher quality of service as an option for applications and content providers such as Google, Amazon, Vonage, and Yahoo to reach end users. In general, Internet companies are opposed to the idea of having to pay access fees to reach end users and support legislation to mandate the existing neutral Internet for they believe that neutrality is most conducive to enhancing the incentives of applications and content providers to innovate. (See Frischmann and van Schewick, in press, for a compelling argument that fostering innovation in Internet applications and content is the primary motivation behind net neutrality legislation). Proponents of net neutrality desire a regulatory/legislative safeguard against potential discriminatory behavior, as manifested in a two-tier Internet, by broadband service providers toward unaffiliated applications and content providers.

Opponents of network neutrality legislation/regulation see this rationale as self-serving and portray legislation as an attempt to reduce competition among existing Internet companies and to create a barrier to entry to applications and content markets (see Sidak, 2007, for a forceful argument for this position). Moreover, critics point out that independent applications and content providers have market alternatives to broadband service providers to improve the quality of their delivered service. Two examples are to self-provide distributed capacity and to purchase enhanced quality of service from competitive content delivery network providers such as Akamai.

One competition issue associated with network neutrality examines the state of horizontal competition between broadband service providers. From a customer's perspective wishing to purchase a broadband connection to the Internet, net neutrality proponents see very little structural competition (most consumers face a duopoly market structure) while net neutrality opponents see strong behavioral competition (declining prices for broadband Internet access and better quality connections) between telephone and cable companies, as discussed above, with potential competitors on the brink of entry into the market. Opponents could cite as an example of potential competition the announcement by Google to purchase spectrum in a forthcoming FCC auction to build purportedly a national mobile carrier network. This network could serve as an open platform to the Internet (Delaney and Sharma, 2007). Regulatory decisions not to require open access to the physical layer owned by wireline broadband access providers implies that structural competition to the incumbents will come from new technologies, such as wireless, and not from intra-platform (intra-modal) wireline competition. (European regulators continue to see the benefit to society from fostering intra-platform competition by proposing to separate the incumbent wireline provider's network into wholesale and retail divisions, referred to as "functional separation." For a criticism of this approach, see Crandall, 2007.) Moreover, some critics of net neutrality legislation argue that policy makers should not focus their efforts on promoting competition in the applications and content layers but rather in the physical layer for that is the location of concentrated economic power in the network layers model. (See Yoo, 2005, for an articulate argument for this position.)

As the market reaches a mature state, a related issue impacting consumer welfare in the broadband access market concerns the significance of switching costs. Future competition will involve attempts to entice consumers to switch broadband technologies as opposed to convincing consumers to upgrade to a broadband connection. To the extent that economic and technical obstacles hamper the ability to switch easily technologies for broadband access, the market power of incumbent access providers is enhanced, even in a market with several rival broadband suppliers. A militating factor to broadband access providers' ability to exercise economic power over residential customers is the availability of commercial and institutional broadband access for consumers when they are at work or school. Such access would serve as a substitute to broadband access at home (Owen and Rosston, 2003).

A two-tier Internet, consisting of a private/managed Internet and a public Internet, can give rise to an exception to the Chicago-School theory that broadband service providers will only encourage efficient use or discourage inefficient use of its platform. The potential for anticompetitive vertical strategic behavior by incumbent broadband access providers exists to the extent that services of unaffiliated applications and content suppliers are viewed as good substitutes or potential substitutes for Internet services telephone companies actually offer or plan on provisioning. The most obvious example of the incentive for a telephone company to discriminate in an anticompetitive manner is toward an Internet voice competitor such as Vonage. For example, a telephone company could block access to its broadband customers for a close substitute VoIP service. Such straightforward and transparent discrimination is likely to lead to a political and regulatory backlash. (This happened in 2005 when the FCC "fined" a rural telephone company for blocking access to its broadband customers to a VoIP provider's service.) More likely, telephone companies may confine competitive substitute services to the public Internet in which the quality of service will often be inferior in times of natural or artificially-created congestion to that provided by the telephone company's service over its circuit-switched telephone network or the private Internet it reserves for its own or affiliated companies' services.

As telephone networks continue to converge to the Internet network's four layers model, telephone companies will increase their ability to provide upper layer services on top of their broadband platform and thus expand the Internet services that they view as good substitutes. (This is similar to the incentive that platform operator Microsoft possesses to expand the features and functionality of its operating system by encompassing services previously offered by independent applications providers.) A good example of this is telephone companies' provision of IPTV over a private/managed Internet. Consumers are increasingly looking to the public Internet for access to television programming and amateur videos. The accompanying increase in demand for Internet television viewing is bandwidth-intensive and thus taxing the capacity and capability of the best efforts Internet to handle all of the varying demands on its resources. Pioneers of the original Internet argue that the public Internet will soon face a crisis situation as bandwidth intensive applications increase in demand (White, 2007). It is likely that congestion on the public Internet will often lead to inferior video and latency-sensitive experiences for consumers and make the private/managed Internet video offerings of telephone companies appear relatively more attractive. In a sense, congested public channels will translate into an advantage for integrated and affiliated broadband access providers' services by acting as a form of raising independent rivals' costs. Thus, to the extent that access to the scarce channels on the private/managed Internet will grow in demand, this will result in increasing revenues attributably to the private/managed Internet.

Broadband service providers possess an incentive to discriminate against substitute services by denying them access to these managed channels while promising that the public Internet will remain open to competitors. Such an incentive could distort competition in Internet markets that require a higher quality of service than the overtaxed public Internet can provide. This could result in Internet service markets (such as the delivery of video programming/content) becoming concentrated in structure due to the network design and investment decisions by broadband service providers. A perverse incentive exists for broadband service providers since as they increase investment in the public Internet to improve its capacity and performance, the return on their investment in the private/managed Internet decreases. If these events transpire, it is predicted that the next round of the net neutrality debate will involve heated demands for mandated open access to the broadband service providers' private/managed Internet channels (Faulhaber, 2007, p. 699).

IV. CONCLUSION

Telecommunications companies are well on their way to transitioning circuit-switched dedicated voice networks to packet-switched broadband networks that can support a wide range of IP-based applications and content. Responding to this inexorable path to technological convergence is a federal telecommunications regulatory scheme that has moved from reliance on common carrier regulations to control the market power of telecommunications companies to trust in competition and increased investment in broadband technology to promote consumer welfare. Both convergence and deregulation have positioned telecommunications companies to be strong competitors in this new Internet-based world.

It is likely that the carriers will continue to integrate vertically into applications and content markets (layers 3 and 4 in the network layers model) and exploit any competitive advantages associated with a two-tier Internet. Telecommunications companies are familiar with a network in which they control the intelligence, suggesting that they are likely to feel more comfortable with a two-tiered Internet that includes at least one intelligent tier. Limited capacity available in the intelligent private/managed Internet promises to reduce structural competition for some services, such as video programming distribution.

Telecommunications companies will counter criticism of this strategy by pointing to the preservation of neutrality on the increasingly congested public Internet. In this way, incumbent broadband access providers continue to encourage consumer demand due to the wide variety of services they provide open access to, while, at the same time, position their integrated and affiliated services to perform better for the consumer. The tiered position of applications and content is one factor that can contribute to the success or failure of a service. An interesting parallel has been uncovered in the cable television industry. Chen and Waterman (2007), in a recent empirical study, find that "…integrated systems that do carry rival networks often position them on digital tiers having more limited subscriber access" (p. 249). This puts independent cable networks at a competitive disadvantage, notwithstanding the program access rules. Similarly, telecommunications companies will continue to open their networks to rival services but possibly only on the public Internet. It is, thus, likely that a two-tiered Internet will (1) dampen the incentive to create new applications and content, knowing that one is confined to an increasingly congested distribution network, and (2) encourage independent applications and content providers to seek intensively for alternative access to consumers. (Superior physical and logical infrastructure can be considered a good substitute for non-neutral quality of service protocols in providing faster, more reliable transmission of data.) One theme of this research is that the architecture of the Internet is a contributing factor in determining the competitiveness of Internet services especially for services that require intense bandwidth or steady transmission speeds. Moreover, once the architecture is in place, changes to that architecture will take significant time and investment.

But could it be that the advantages of a two-tier Internet are fleeting? If incumbent broadband access providers along with new entrants continue to upgrade the performance of the public Internet, the competitive advantage of a two-tier Internet will be eroded over time. Public policy makers' best bet for such an outcome should focus on fostering new entry by increasing the opportunities to generate revenue and/or reducing the costs of entry into the physical layer.

REFERENCES

Barris, M. (2007). Cablevision Posts Wider Loss, Tempers Outlook for New Services. *The Wall Street Journal Online Edition*, November 8.

Boles, C. (2007). AT&T Wins Partial Relief from FCC On Some Broadband Regulations. *The Wall Street Journal Online Edition*, Retrieved October 12, 2007.

Britton, B. B., & McGonegal, S. (2007) *The Digital Economy Fact Book, Ninth Edition 2007* (p. 61).Washington, D.C.: The Progress & Freedom Foundation.

Brumfield, C. (2007a). Time Warner Cable Falls to 50% Penetration Level. *IP Democracy*, November 8, 2007, Retrieved May 8, 2008 from http://www.ipdemocracy.com/archives/002755time_warner_cable_falls_to_50_penetration_level.php.

Brumfield, C. (2007b). Is Cable Telephony Growth Tapped Out? *IP Democracy*, November 15. Retrieved May 8, 2008 from http://www.ipdemocracy.com/archives/002765is_cable_telephony_growth_tapped_out.php.

Chen, D., & Waterman, D. (2007). Vertical Ownership, Program Network Carriage, and Tier Positioning in Cable Television: An Empirical Study. *Review of Industrial Organization*, 30(3), 227-251.

Crandall, R. (2007). Technical Difficulties. *The Wall Street Journal Online Edition*, November 8, 2007. Retrieved November 8, 2007.

Delaney, K., & Sharma, A. (2007). Google Has Even Bigger Plans for Mobile Phones. *The Wall*

Street Journal Online Edition, November 16, 2007. Retrieved November 16, 2007.

Farrell, J., & Weiser, P. (2003). Modularity, Vertical Integration, and Open Access Policies: Toward A Convergence of Antitrust And Regulation In The Internet Age. *Harvard Journal of Law & Technology, 17*(1), 85-134.

Faulhaber, G. (2007). Network Neutrality: The Debate Evolves. *International Journal of Communications, 1*(1), 680-700.

Federal Communications Commission (2005). *Appropriate Framework for Broadband Access to the Internet over Wireline Facilities.* Policy Statement, CC Docket No. 02-33, FCC 05-151.

Federal Communications Commission (2007). *High-Speed Services for Internet Access: Status as of December 31, 2006*. Industry Analysis and Technology Division, Wireline Competition Bureau, October 2007.

Frischmann, B., & van Schewick, B. (in press). Network Neutrality and the Economics of an Information Superhighway: A Reply to Professor Yoo. *Jurimetrics, the Journal of Law, Science and Technology.*

Grant, P. (2006). Cable Industry May need to Spend Heavily on Broadband Upgrades. *The Wall Street Journal,* August 17. p. A1.

Malik, O. (2007). 3Q 2007 Broadband Report Card: B-Minus. *GigaOm,* November 13. Retrieved May 8, 2008 from http://gigaom.com/2007/11/13/3q-2007-broadband-report-card-b-minus/.

Mangalindan, M., & Rhoads, C. (2007). Sorry, Wrong Number, eBay Says on Skype. *The Wall Street Journal,* October 2, p. B1.

Moritz, S. (2007). AT&T, EchoStar Deal Nears. *TheStreet.com,* November 19.

Nuechterlein, J., & Weiser, P. (2005). *Digital Crossroads.* Cambridge, MA: The MIT Press.

Owen, B., & Rosston, G. (2003). Local Broadband Access: *Primum Non Nocere* or *Primum Processi?* A Property Rights Approach. *Joint Center AEI-Brookings Joint Center for Regulatory Studies, 03*-19, p.7.

Peha, J., Lehr, W., & Wilkie, S. (2007). The State of the Debate on Network Neutrality. *International Journal of Communications, 1*(1), 709-716.

Picker, R. (2005). SBC and Verizon: Judgment Day. *The University of Chicago Law School Faculty Blog,* November 3. Retrieved November 3, 2005 from http://uchicagolaw.typepad.com/faculty/2005/11/sbc_and_verizon.html.

Reardon, M. (2006a). Qwest CEO supports tiered Internet. *c/net NEWS.Com*, March 16, 2006. Retrieved March 16, 2006 from http://news.com.com/Qwest+CEO+supports+tiered+Internet/2100-1034_3-6050109.html.

Reardon, M. (2006b). FCC approves AT&T-BellSouth merger. *c/net NEWS.com*, December 29, 2006. Retrieved May 15, 2008 from http://www.news.com/FCC-approves-ATT-BellSouth-merger/2100-1036_3-6146369.html.

Schatz, A. (2007). FCC Deregulates Verizon's Big-Business Market. *The Wall Street Journal Online Edition*, Retrieved March 21, 2007.

Searcey, D. (2007). Cable's Picture Gets Fuzzier. *The Wall Street Journal*, November 8, 2007. p. B3.

Shwiff, K. (2007). Legal Costs Widen Vonage's Loss. *The Wall Street Journal Online Edition,* Retrieved November 8, 2007.

Sidak, G. (2007). What Is the Network Neutrality Debate Really About? *International Journal of Communications*, *1*(1), 377-388.

Swanson, B. (2007). The Coming Exaflood. *The Wall Street Journal Online Edition*, Retrieved January 20, 2007.

Verizon (2007). Verizon Continues to Dramatically Raise Broadband Upload Speeds in FiOS Internet Service Area. Verizon News Release, November 20.

Weiser, P. (2003). Toward a Next Generation Regulatory Strategy. *Loyola University Chicago Law Journal,* 35, 41-85.

Werbach, K. (2002). A Layered Model for Internet Policy. *Journal of Telecommunications & High-Technology Law, 1*(37), 58-64.

White, B. (2007). Its Creators Call Internet Outdated Offer Remedies. *The Wall Street Journal Online Edition*, October 2, 2007. Retrieved October 2, 2007.

Whitt, R. (2004). A Horizontal Leap Forward: Formulating a New Communications Public Policy Framework Based on the Network Layers Model. *Federal Communications Law Journal, 56*(3), 587-672.

Yoo, C. (2005). Beyond Network Neutrality. *Harvard Journal of Law & Technology, 19*(1), 1-77.

KEY TERMS

Convergence: The blending of three previously-distinct services: voice telecommunications, video (television) service and Internet access.

Federal Communications Commission: The United States federal regulatory body that oversees competition issues in the Internet service, telephone and cable television industries.

Net Neutrality: The decision by the Internet service provider to treat all content and applications equally in terms of distribution to the end user.

Network-Layers Model: A model of a communications network that illustrates the distinction between content, applications, software and physical infrastructure.

Telecommunications Act of 1996: A federal United States law that governs the regulation of telephone and cable companies.

Triple Play Provider: A company that provides all three important communication services to a household: voice communication, video (television) and Internet access.

Two-Tiered Internet: The decision by the Internet service provider to give some content or applications preferential treatment, for example by using faster transmission speeds.

Chapter IX
International Internet Interconnection Service in Asia-Pacific Region

Moon-Soo Kim
Hankuk University of Foreign Studies, Korea

ABSTRACT

Since 1986, the Internet has developed into a global network enabling users worldwide to connect to each other to exchange information and data. The initial zero settlement peering arrangements, however, have now largely been replaced by commercial transit arrangements, as backbone providers seek to recoup their network infrastructure investments and generate commercial profits. This is a key cause of the issues and debates that have emerged between developed and developing countries about international Internet interconnection services (IIS). This study focuses on current interconnection settlement arrangement models that disfavor ISPs and end-users in the Asia-Pacific region. After reviewing the Internet market and digital divide in the region, the chapter summarizes the main current IIS issues between the Asia-Pacific and Western regions into three categories of concern: inequity, anticompetitive practices and the threat of the "balkanization" of the Internet. Practical recommendations to resolve these issues and improve the Asia-Pacific IIS market are discussed from regional and international perspectives.

1. INTRODUCTION

The Internet is commonly understood to be a worldwide network reaching beyond national borders. This worldwide connectivity, however, does not mean that there is one global network serving all regions across the world. Instead, this connectivity is created by countless open-architecture networks of varying sizes, interconnecting to exchange information and data. The original free peering arrangement, however, has now been replaced by transit arrangements, as backbone providers have sought to recoup their network infrastructure investments and generate commercial profits. Transit arrangements constitute a special problem for ISPs in the Asia-Pacific region with regard to their interconnection to American IBPs (or 'backbones') and ISPs from other regions, as such Asia-Pacific/US interconnections are provided under one-sided transit arrangements, which entail the full costs of such connections being borne by the Asia-Pacific region.

This type of unilateral settlement arrangement seemed to make sense up until the 1990s, when most internet-related information was concentrated in the US, but this is no longer the case. ISPs in Korea and other Asia-Pacific countries have been paying excessive prices to IBPs in the US and other developed countries on the basis of such agreements, a situation that is exacerbating the digital divide between world regions. The Asia-Pacific countries officially challenged this inequitable cost-sharing arrangement at the third APEC TEL Working Group meeting in Singapore in 1998. During this meeting, a special taskforce (the International Charging Arrangement for Internet Service; ICAIS) was formed, and a resolution adopted to conduct research in this area[1]. However, investigating the internet connection market is far from easy, as details of interconnection arrangements are considered as purely commercial agreements which have been freely consented to by the parties involved, and are thus generally kept confidential. Typically, such contracts impose confidentiality not just as to their principal content, but also about pricing details and even the most general information. Challenges to research investigations also come from the sheer complexity of internet traffic, which makes it more difficult to design rational settlement models than, for example, for telephone networks. Efforts to resolve the imbalances in charging arrangement have thus so far failed to produce concrete results.

This chapter, as a pilot study under the deficiency of the related data and previous studies, deals with the IIS (International Internet Interconnection Service) market, particularly, the current interconnection settlement arrangement which disfavors ISPs in the Asia-Pacific region, further is to give the alternative methods to resolve this issue by examining the debates among internet-related experts and policymakers, reviewing other records containing the content of meetings of international organizations such as WTO, APEC, OECD and ITU-T, etc.

The chapter is arranged as follows. Section 2 offers the various types of arrangements in its hierarchy of interconnection. Section 3 reviews the recent trends in the Asia-Pacific internet market from the perspective of the digital divide between the developed and developing countries, while section 4 examines several issues raised in recent years about interconnection services between the Asia-Pacific and Western regions. Practical recommendations designed to resolve these issues and to develop the Asia-Pacific IIS market are discussed in section 5. Finally, section 6 includes the summary and concluding remarks.

2. IIS AND INTERCONNECTION AGREEMENTS

The Internet is a network of computer networks operated by various ISPs using a universal protocol known as TCP/IP. Traffic exchange occurs both between ISPs located in different service geographic regions, and between those at different tiers of the internet connection hierarchy. Arrangements for physical interconnections between these ISPs may be distinguished into two broad categories: peering arrangements – which involves the cost-free transfer of data - and transit arrangements, where ISP s or IBP's charge for such transfers. Thus internet connectivity is ensured via one or more of these peering and/or transit arrangements.

IBPs forward traffic generated by both their own customers and subscribers to other ISPs on to outside regions, interconnecting with other IBPs to exchange data under connection agreements referred to as peering contracts. Peering arrangements generally occur between Autonomous Systems (ASs), usually held by IBP or big ISPs of equivalent size, or at least where both parties can see a mutual benefit (OECD, 2006). In reality, there are 3-types of peering: public peering, private peering and paid peering.

Private and public peering are distinguished by the degree of openness of their interconnections, but generally involve free peering, an arrangement usually adopted when the two parties consider the amount of traffic from each sides to be roughly equal, thus making a settlement-free exchange more cost-efficient. Free peering is most often entered into between two ISPs of a similar size, and the arrangement is known as "bill-and-keep", "settlement-free" or "sender-keeps-all" (this does not include special cases like paid peering).

Public peering occurs at public access points or exchanges such as NAPs or Metropolitan Area Exchanges (MAEs) in the US. At these public access points, large numbers of networks exchange traffic using the so-called "hot-potato routing" method. Public peering is managed by public organizations, but as internet traffic volumes increased, NAPs serving as public peering points were affected by heavy congestion, and *private peering* emerged as an alternative to public peering. Private peering consists of directly connecting circuits between two large IBPs. Under private peering, there is no risk of information disclosure, as the exchange of traffic takes place on agreed terms at access points agreed between the ISPs. Private peering constitutes a crucial area for global internet and related traffic issues, providing as much as 80% of total US internet traffic (Elixmann & Scanlan, 2002).

Lastly, *Paid peering* is used when two ASs (Autonomous Systems) agree that, on balance, one of the networks derives greater value from traffic exchange than the other. For example, ISP-A may have a lot of customers or attractive content, and ISP-B agrees to pay ISP-A so that its customers may have direct access to ISP-A's customers and content, although ISP-A will not carry third party traffic for ISP-B. A network such as the AOL Transit Data Network (ADTN) offers paid peering to other networks, allowing their customers the benefits of direct access to AOL Time Warner content as well as to their 35 million customers (OECD, 2006).

If ISPs are unable to enter into peering agreements with other IBPs, they have to buy transit services from an IBP in order to provide their subscribers with universal access to the global internet. *Transit* is the term applied when one AS agrees to carry traffic for another for delivery to the rest of the Internet. Whereas peering only offers connectivity between the customers and content of two individual ASs, transit usually provides connectivity to the entire internet at a predictable price. Transit providers such as global IBPs, regional IBPs and big ISPs, etc charge for both basic and ancillary services, such as Service Level agreements, installation support, local telecommunication infrastructure provisioning and NOC (Network Operations Center) support.

Transit arrangements are the most common interconnection model, and are vital in guaranteeing the Internet its global connectivity. Under such arrangements, an ISP wanting to send traffic to another ISP which has to be routed via a network owned by a third ISP must purchase a corresponding transit service. An ISP (or IBP) that sells transit services will receive packets that are not addressed to it as their final destination, and route them toward other ISPs with which they have peering relationships or, where they lack such arrangements, buy transit services to forward them. Transit can provide end-users of the Internet with a degree of connectivity which could never be possible via peering alone, and thereby enable ISPs to offer many more services. In sum, ISPs enable user access to worldwide internet address locations through a combination of peering and paid transit arrangements. When new IBPs providing transit services enter the market, they trigger price competition similar to that in any other market, which may bring down transit fees, perhaps to the level of connection fees. Meanwhile, from an IBP's point of view, connecting with more ISPs means increasing their network effect, which also gives them an incentive to seek to lower transit fees.

3. THE ASIA-PACIFIC IIS MARKET AND THE DIGITAL DIVIDE

The Asia-Pacific region, home to 56.3% of the world population, accounted for about 38% of all internet users worldwide in 2004, a share which was projected to grow to 42% by 2007. This will make it the regional internet market with the largest number of subscribers and which generates the most traffic, as well as being the source of the biggest internet investment and offering the highest value-added potential. However the number of internet subscribers in the Asia-Pacific region amounts to only 8.3% of total regional population[2] in 2003, and there exists significant disparities of subscription rates between Asia-Pacific countries. This gap is explained by the difference in national internet infrastructures, internet investment conditions and national levels of interest in using the Internet. In 2003, the three wealthiest of the 40 Asia-Pacific countries accounted for 28% of total internet subscribers in the region; with the nine countries in the medium-developed or upper-income country group accounting for 16%; and the remaining 28 countries for the final 56% (ITU-T, 2004).

The disparity between Asia-Pacific countries is even wider in terms of broadband internet subscribers, where the 12 wealthiest Asia-Pacific countries account for an overwhelming 84% of the region's broadband subscribers. These extreme imbalances in terms of internet usage are the main causes of the digital divide that exists between the region's countries. Meanwhile, in terms of international internet traffic levels, the medium-developed or upper-income countries with a large number of broadband subscribers occupy a 58% share of total traffic. Also, the international internet traffic amount per capita among the 28 lower-income countries represent to less than 1% of the average. The remaining 12 countries account for over 99% of international

Figure 1[3]. Shares of Internet and broadband internet subscribers by country Group in Asia-Pacific Region

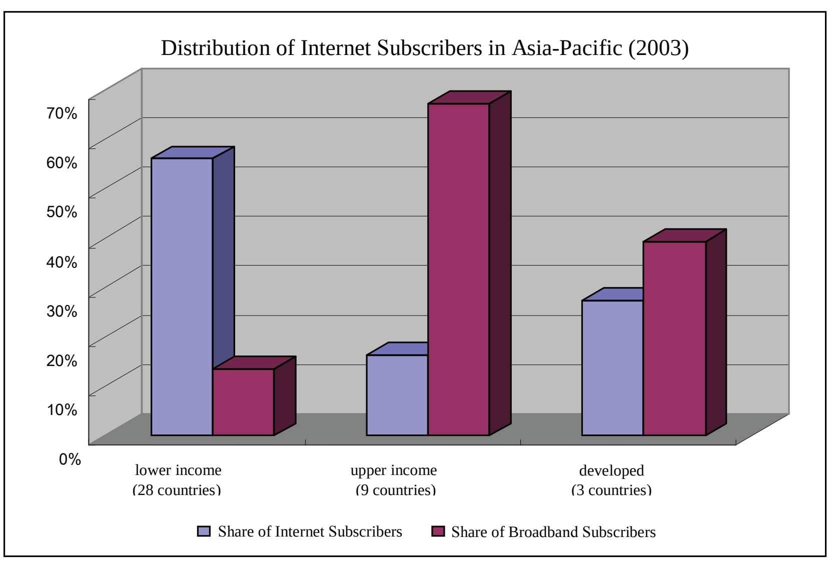
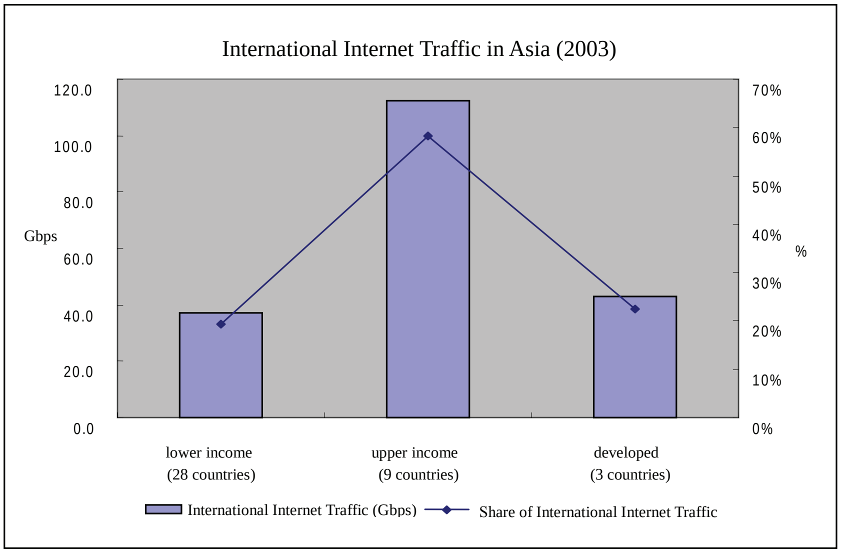


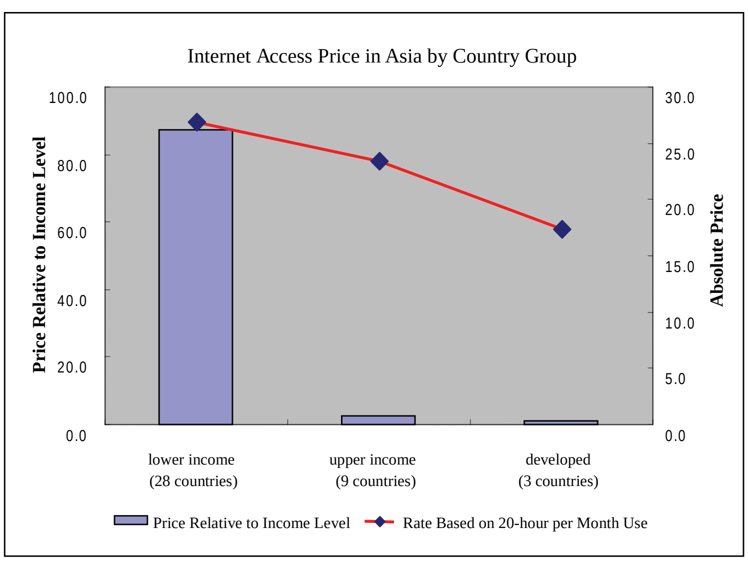
Figure 2. Internet Traffic Amount by Country Group in Asia-Pacific Region



internet traffic, indicating the lopsidedness of the region's internet usage.

The chief cause of this considerable disparity in internet use within the region appears to be the high internet access price charged to end users, which is driven up by the absence of internet infrastructure and high price of international connections. In 2003, the average monthly internet access fee (based on 20 hours use per month) in the 28 lower-income Asian countries was 1.6 times that pertaining in the three wealthiest countries, which translates to being 88 times more expensive when these countries' per capita income levels are factored in. There was also a sizeable difference between upper-income and lower-income countries in terms of internet access service pricing structures. In the former, flat monthly rates without set usage limits were the

Figure 3. Internet access price by country group in Asia-Pacific Region



norm, whereas lower-income country users faced separate charges for excess minutes or hours, as well as telephone charges and other additional fees, which obviously hinder local internet users.

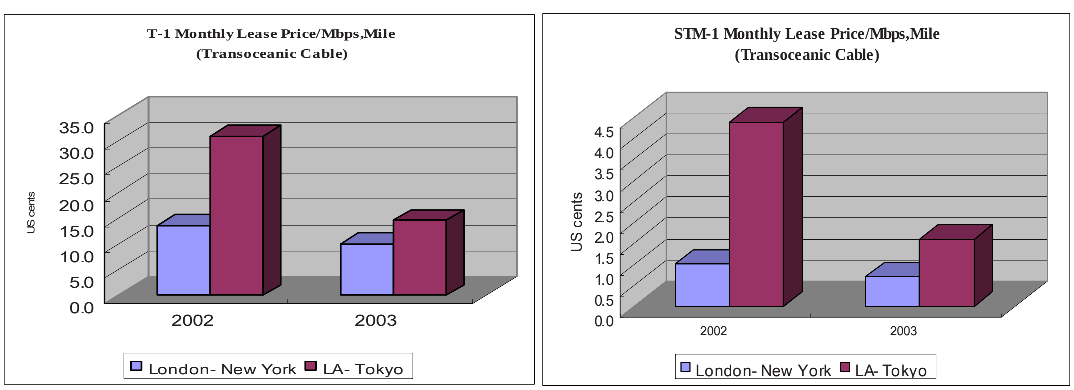
This disparity at the level of the number of internet users and traffic amounts is inevitably echoed in market size and investment. The fact that the capital expenditure shares of China and other Asian countries exceed those of their revenue by far is encouraging for the regional telecom industry. Meanwhile, the size of the internet backbone market of the Asia-Pacific region is nevertheless expected to reach only a size corresponding to 50% of that of Europe and North America by 2007 (RHK Inc. 2003). Notably, the fact that the monthly traffic levels generated in Asia-Pacific during 2004 were about equal to that in Europe[4] suggests that the Asian backbone market is less competitive than its European and North American counterparts. This inequality in turnover, on the other hand, explains why Asian backbone providers have continued to have to purchase transit services from European or North American providers, rather than being able to enter into peering relationships as equal partners. Meanwhile, the European backbone market is expected to roughly equal the size of its North American counterpart, evolving along a different path than the Asian market, despite the later generating about the same amount of traffic.

Figure 4[5] shows that, while international interconnection prices have declined overall, the costs paid by Asia-Pacific ISPs are much higher than their European and US counterparts. Interconnection charges (referred to as upstream bandwidth cost, see AfriISPA, 2002) make up the largest element of most ISPs' operating costs, as well as a considerable portion of the internet access fees paid by end-users in emerging Asia-Pacific countries. According to Antelope (2001) the share of ISP costs accounted for by international internet transit charge, which (while varying quite widely from country to country) was estimated in 2001 as approx. 30% on average.

This considerable cost burden is a main factor behind Asian complaints as to the inequity and unfairness of their current international transit agreements with US and European IBPs. If the digital divide among Asia-Pacific countries is to be reduced by increasing investment and promoting diffusion of internet accessibility, it is seen as crucial that ways are found to lower the international transit costs that make up such a large portion of the region's ISP's operating budgets.

This analysis of internet trends in the Asia-Pacific region highlights the following general points. First, the rapid increase in internet users' numbers and traffic levels has only worsened the problem of disparity between the region's countries, and this gap is even wider in the broadband

Figure 4. International Internet lease price, US cents/Mbps,Mile



internet service market. The imbalance can also be observed in backbone connectivity, which is concentrated in a handful of countries in the region. Second, the main reason behind this digital divide appears to be the internet access prices charged to emerging countries, which are significantly higher than in developed countries. Third, one of the major components of this high internet access price is the costs of international connection, for which Asian countries pay about twice what European or North American ISPs are charged for transatlantic links. Fourth, considering how the volume of intra-regional traffic destined for Asia-Pacific locations is rising fast, it is vital that Asia-Pacific countries cooperate to find solutions to bring down the international transit prices substantially.

4. IIS MARKET AND AGREEMENT ISSUES

This review of the IIS market in Asia-Pacific highlights the tension and discord surrounding international internet interconnection services, which divide the global internet industry into two camps: North America and Europe on the one hand, and Asia-Pacific and the rest of the world on the other. The arguments against the IIS model hinge on two questions; whether the existing interconnection service arrangements are equitable, and whether the behaviors of international IBPs in the IIC market can be considered anticompetitive. This controversy between the two halves of the world is stoking fears about the eventual 'balkanization' (or fragmentation) of the Internet, and anxieties are being voiced that this discord might flare up into something irreversibly damaging, putting an end to the universal connectivity of the Internet as we know it today. It is our view that these three issues (inequity, anticompetitive behavior and the threat of balkanization) will ultimately necessitate regulatory intervention in the international internet connection market.

4.1 Inequity Issues

The issue of the inequity of the current international connection service model can be discussed under the following four headings. The first point concerns the current hierarchical structure of worldwide internet connection models that reserves peering arrangements to certain ISPs, with others having to purchase transit services. The argument is basically that this hierarchical structure discriminates against small providers and in favor of larger ones, and forces developing countries to pay unilateral transit fees to developed countries. In a brief submitted to ITU-T, China (2002) called for variations to the settlement model, arguing that the high prices paid by ISPs and users in Asia-Pacific and in the emerging world under the current international internet settlement system amount, in effect, to subsidies toward North American IBPs and users, and contribute to worsening the digital divide between world regions. Meanwhile, a US FCC (Federal Communications Commission) study has claimed that there was no evidence of discriminatory behavior by American ISPs in their interconnection agreements with ISPs from other countries, but found that these agreements treated American and foreign backbone operators equally (Kende, 2000). However this response is not credible, for a number of reasons. First, the FCC does not have access to concrete details of IBP agreements[6], and therefore cannot say with certainty that American IBPs have not behaved in an anticompetitive manner. Secondly, even if it had uncovered discriminatory or anticompetitive behavior, the FCC has no power to take any meaningful form of action, since (unlike telephone and other telecommunications services) internet-related services providers are under no obligation to offer transparent, rational and nondiscriminatory conditions to third parties with which they do business. If evidences were found that American IBPs had abused their market dominance, corrective actions would require antitrust procedures to be initiated under US antitrust laws: however, the backbone transit market could be technically considered an overseas market, and thus beyond the scope of these laws. A significant problem is that the US does not perceive international interconnection issues to be a problem: for developing countries, however, it seems unfair that large chunks of their telecom revenues from internet access fees are diverted to overseas backbone operators, reducing their provider revenues and ultimately lessening their financial ability to invest in internet infrastructure and frustrating their governments' telecom revenue diversification policies. (The logic of the developing countries' view on this matter is very similar that of those US telcos that are currently recording losses in their international telephone business, and complaining that US telephone subscribers are effectively subsidizing users in other countries.)

The second point centers on the fact that developing countries that have to pay unilateral transit fees consider it unfair that they have to cover traffic costs both to and from the US by paying connection fees and circuit charges for two-way traffic. Large ISPs can deal with the problem simply by circumventing backbone networks, but only a handful of ISPs in developing countries are large enough to afford to do this. Some have proposed splitting the internet interconnection costs between interested parties, with ISP shares being determined by traffic levels. Objections to this solution have pointed out that other factors must also be taken into account, but there is currently no technological means to reflect these in cost estimation terms. The American position in this matter is that, since all connections to the worldwide internet transit via the US (as has been the case since the early days of the Internet), and the great majority of content is also located there, traffic of US origin occurred in other countries benefits users in these countries rather than the other way around, and that therefore these countries should pay both-way traffic cost. The OECD (2001) supports the American argument, denying that there is any cause for concerns about equity in the current settlement system. Their arguments claim that peering is steadily rising and transit prices declining, and that the increasing replacement of transit connection models by peering will undermine the hierarchical characteristics of the Internet. However, it has also been pointed out that top-tier IBPs are currently reducing the number of ISPs with whom they have settlement-free peering relationships. The position of most developing countries (including China) is that, while they acknowledge that much of the early development of the Internet took place in the US, and do not reject transit arrangements as such, there is a need for a new international interconnection agreement model, as the pattern of global internet traffic flow is no longer so US-centered, a trend that is likely to accelerate further in the future.

The third point is the inequity of charging developing countries high full circuit charges and network fees when a steadily increasing amount of traffic within the Asian or Oceania regions does not actually transit through the US, and while the unit price of circuit connectivity is declining. American IBPs claim there is a lack of convincing evidence or examples to support this case, which further calls into question the reliability of the methods used for analyzing existing data. The technical issues regarding the reliability of the relevant data and examples and their analyses could be resolved relatively easily if efforts were made to identify valid methodologies.

The final point is that the current international internet settlement system is fundamentally designed to benefit the progress of developed countries' internet industries. Developing countries are protesting that the circuit and access costs they pay are financing developed countries' IBPs and end-users, and that this is only serving to widening the digital divide in the world. To give a concrete example, almost 23% of total national expenditure on internet access in Mexico is composed of international connection fees (CITEL, 2001), while, as noted previously, according to Antelop (2001), about 19-80% of ISPs' total costs are international internet connection-related, and that this is also an important component (about 10%) of end-user access prices. This is not negligible, especially for the least developed countries. Moreover, developing countries' ISPs are required to renew their transit contracts with IBPs annually, which adds to their financial burdens. The major component of the developing countries' complaint is that the high fees charged to their end-users hinder the growth of the subscriber bases in these countries, and the resulting modest demand inhibits their internet industries from reaching a sufficient size where they can benefit from economies of scale. Where their ISPs try to reduce these prices, the result is decline in service quality and ultimately the stunting of the growth of these countries' internet industries.

4.2 IBPs' Anti-Competitive Behavior

Representatives of Asia-Pacific ISPs have pointed to six types of anticompetitive practices by international IBPs in the IIC market: absence of transparency in transactions; the domination of the market by existing providers; monopoly pricing and tacit collusion between IBPs; withholding of IP address resources; interference with other ISPs' business operation; and the use of non-transparent criteria for determining which providers are eligible for peering arrangements.

The complaint about lack of transparency in IIC agreements is directed against the behavior of a number of international IBPs who, even if they make their corporate policies concerning peering and transiting arrangements public, maintain absolute secrecy regarding the concrete details of these arrangements, and treat them as internal business secrets (Australia, 2002). IBPs also exclude small and medium-size ISPs from technical dialogues regarding issues related to interconnection (Antelope, 2001). Thus ISPs leasing submarine cables or purchasing IRUs (Indefeasible Right of Use) cannot participate in submarine cable consortiums' decision-making, which is surprising, given their need for detailed information (including technical standards data) about circuits they connect to. Types of anti-competitive behaviors practiced by IBPs include 'refusal to deal,' 'price squeeze' (causing cost increases for competitors), 'predatory pricing,' 'refusal to deal in certain services or certain regions,' 'tying arrangements with long-haul backbone trunks, etc' and 'forcing the purchase of unwanted services' (Roseman, 2003). The absence of transparency caused by the nondisclosure of peering or transit agreement terms (and other forms of IBPs anticompetitive conduct) could be dealt with through regulatory efforts. However, proponents of this argument have so far been unable to thoroughly demonstrate its credibility due to the lack of supporting evidence, concrete examples or studies on the subject.

Anticompetitive domination practices exhibited by existing IBPs in the IIC market have two major characteristics. Where control over essential facility provisions is limited to one or a few suppliers, who cannot easily be substituted for economic or technological reasons or because of their market position, such suppliers are free from competition and can adjust prices or control supply as they wish. A supplier meeting either of these criteria is usually perceived as a major supplier[7] to the relevant market. Clearly, the greater the market share of a dominant IBP, the more small/medium ISPs will have to obtain transit services from them. There are several major international internet backbone suppliers in this position, and their market dominance is a major concern.

Tacit collusion is widely assumed to be taking place in this market, raising strong suspicions that IBPs are practicing monopoly pricing. Australia (2002) considers that top tier backbone providers are engaging in collusions of various types including such typical cartel behaviors as refusing to offer usage-based agreements and demanding that ISPs connect to certain networks in specific geographical locations. Also China (2002) asserts (without actually using the word collusion), that pricing in IIS markets is anti-competitive. Again, the allegations of abuse of market power by price fixing, collusion and other anti-competitive practices are weakened by the lack of concrete proofs or examples. IBPs maintain that backbone markets worldwide (including in the US and Europe), are competitive markets, and influential international organizations such as the OECD and EC have lent their support to this argument.

IBPs are also accused of abusing their market position in the area of IP address resources. The control vast of IP address holdings by IBPs is seen as an anti-competitive practice, since only much smaller shares are available for small ISPs or those from developing countries, contributing to their weak competitive positions. This argument is countered by those who point out that there are still many unclaimed IPv4 address resources, as well as virtual addresses which can be used instead, and that, even if prices are affected by the cost of migrating to IPv6 addresses (which offer nearly infinite resources), such a transition will take place progressively, and therefore the IP address issue does not have a significant anti-competitive effect on the market.

The fifth type of anticompetitive practice concerns operational matters to do with actual international internet connections. The argument is that IBPs can lower interconnection service quality or delay desirable connection capacity increases at will.

The criteria used by IBPs for selecting peering partners are also alleged to be an anticompetitive market practice. Determining who should be allowed to enjoy peering arrangements and who not is based on the mutual value that can be gained from such a connection by the operators and users involved. However, unlike small ISPs, the big suppliers possessing large networks enjoy considerable bargaining power, as they not only offer access to their own customer base and content holding, but also serve as intermediaries for connecting to other internet networks. This scope of this power elicits concerns about anti-competitiveness.

The Asia-Pacific countries argue that these inequitable practices by IBPs are harming the competitiveness of the IIC market, and that national regulatory bodies or international organizations must intervene to implement corrective actions. Several Asian countries and Australia have joined together to demand the creation of new international regulatory instruments for this purpose: but no progress has been made, and the proposal is in limbo, having encountered several strong objections from established interests.

4.3 The Threat of Internet ‘Balkanization’

The danger of the ‘balkanization’ of the Internet, leading to the destruction of universal internet connectivity, is one of the most important arguments in favor of regulatory intervention in interconnection negotiations between IBPs and ISPs. However, this argument has not so far been brought up in support of their case by those Asia-Pacific ISPs demanding regulatory intervention to change current interconnection arrangement practices. But such a situation could occur if large numbers of small/medium ISPs were to become unable to connect fully to all parts of the Internet.

The phenomenon could result from new business strategies or services launched by IBPs or large ISPs. Industry insiders are projecting that, as future competitiveness strategies, increasing numbers of ISPs will offer new or premium services with guaranteed quality exclusively to their subscribers, but not to those of their peering or transit agreement partners (Elixmann and Scanlan, 2002). The balkanization of the Internet can be considered as more likely to occur should a small number of IBPs come to dominate the market. While web access and e-mail facilities, perceived as universal services, are unlikely to be included in such service differentiation strategies, any ISPs developing new, high-quality services to gain a competitive edge would probably feel reluctant to share them with other ISPs and their customers. There is a strong likelihood that such superior services will be offered using different standards, and customers of regular ISPs or IBPs would not be authorized to access them: if this prognosis becomes the reality, it will obviously pose a serious threat to the Internet's world-wide connectivity.

The introduction of such new high-quality services will tempt top-tier IBPs to discontinue the current style of interconnections that offer 'one-hop' global internet connectivity which links all regions and all content across the world, a development that, again, would further expand their market power. This will lead to a typical network externality effect, where one individual user's choice of a particular backbone network affects the choices of other users. Furthermore, obtaining access to particular content will require connection to a specific backbone network, while content providers wanting to connect to end users would need to transit through the backbone; as a result, the backbone operator's infrastructure will indeed become an essential facility.

The possibility of introducing such differentiation in internet service quality will give IBPs even greater ability to control network interconnections, in turn increasing the possibility of internet balkanization. End-users in developed countries will be able to choose from several IBPs offering differentiated service qualities, but users in developing countries will be left with significantly narrower choices, and ISPs from these countries will be forced to continue to accept exclusionary interconnection agreements. If powerful IBPs introduce service differentiation using different technical standards, this will entail extra software and equipment costs for smaller ISPs, the burden of which will have to be shifted onto their users' interconnection prices. Moreover, such a development would cause peering not backed up by transmission quality guarantees to lose ground in favor of transit arrangements. However, even if this situation actually occurs, it is not clear that it would definitely lead to unambiguously anticompetitive practices.

5. DISCUSSION: ASIA-PACIFIC IIS MARKET

5.1 The Regional Perspective on the IIS Market

The heart of the international internet interconnection issue is the digital divide between developed and developing countries, which is either being worsened, or at least not improved, by the permanent high internet access prices charged to the latters' end-users. Korean internet access fees are not significantly higher than in other developed countries, but the pressure from small/medium ISPs to lower prices is mounting, and Korea's internet connection service market is approaching a saturation point. Korea needs to reduce its inter-ISP connection costs by stimulating competition within its backbone market. Meanwhile, to bridge the digital divide in the Asia-Pacific region by promoting increased internet use, the region's developing countries must launch active efforts to lower not only international internet connection costs, but also the other elements of their overall connection costs, such as intra-national connection expenses (largely telecom network connection costs). The economic and policy-level solution to this problem is the introduction of increased communications market competition, including in the internet interconnection market.

Most countries with developed internet markets feel extremely antipathetic towards the idea of them being regulated. This attitude, which stems from the Internet's history of development and growth as a self-regulated market, is shared by policymakers in most developed countries and by large IBPs. On the other hand, from the perspective of developing countries without big international IBPs, regulation is badly needed, as the current state of the interconnection market negatively affects their ability to provide universal connectivity and to support their end-users' interests. However, regulatory intervention must be preceded by careful consideration of the possible ramifications; for instance, regulatory measures as applied to the telecom industry might turn out to be detrimental to the Internet and jeopardize its growth prospects. A safe and effective option would be to require further transparency in domestic interconnection agreements, and to establish a cost-based accounting rule for domestic IBPs, reflecting actual network costs. In this regard, the regulatory actions recently undertaken by the Korean government for its internet connection market would be a good example to follow (MIC, 2005).

Negotiations about internet interconnection arrangements between providers are commercial in nature, rather than being based on inter-governmental or international organization agreements, and their outcome therefore ultimately depends on the balance of the parties' bargaining power. In the past, most elements of such arrangements, including connection modes (peering or unilateral transit connections) and even settlement methods, were decided by developed country firms: not surprisingly, the outcomes tended to advantage them. Given recent internet trends and international traffic flow changes, there is a good chance that Asia-Pacific will soon see its negotiating power leveraged upwards. On the other hand, there is no real likelihood that the current balance of negotiating power between large IBPs and small/medium ISPs in one-on-one negotiation will alter. Accordingly, rather than seeking to leverage negotiating power at the individual ISPs level, it would be wise to try to use collective bargaining power by setting up a national or regional mechanism to collect all outbound traffic destined for North American or European locations and transmit it together. Buying traffic in large units through national-level or regional-level collective purchase would be substantially more cost-effective than buying in smaller units, and would help reduce the per-unit cost of transpacific circuits. But, to be able to do this, Asia-Pacific governments must remove or ease regulatory restrictions imposed on their ISPs or IBPs with regard to overseas business operations (cross-border supply of internet service, FDI, local joint ventures, etc.). A close-knit network linking Asia-Pacific countries and ISPs would be invaluable.

However, while Asia-Pacific ISPs are still covering the full costs of circuits and transit, changing the settlement model alone will not help much in practical terms. Connection prices themselves need to be brought down, and this cannot be done without boosting these countries' bargaining power. The best way to leverage their negotiating power to achieve their ultimate goal of reducing connection costs is for the countries in the region to share a common stance and speak with one voice on the IIS issue. To be more concrete, first, Asia-Pacific countries must consider setting up a common connection node for all North America-bound traffic, so that a single node exists between the US and the region. This will require hefty circuit investment, and the prospect of sharing such costs might not seem overly attractive to some countries. However, building a connection node at a North American location would enable Asian ISPs to have direct access to numerous other American ISPs, and working together will give them serious negotiating clout which could lead to the possibility of agreeing peering arrangements with them, or at least buying transit services at much reduced price. Second, Asia-Pacific countries must try to gain equal footing with providers at the top rung of the global internet hierarchy, for instance, by developing content and expanding the scope of the regional network. Third, Asia-Pacific countries must create large intraregional NAPs to provide smooth interconnection within the region, and try to lower the amount of traffic transiting through North American routes: reducing the demand for such connections would certainly enhance their negotiating position. Fourth (coupled with reducing North American traffic) Asia-Pacific ISPs must investigate the possibility of setting up caching, mirroring and CDN hubs to serve the region, which would again help minimize the amount of inbound traffic from North America and decrease transpacific circuit capacity. The absolute prerequisite for all these measures to actually produce positive effects on Asia-Pacific ISPs' bargaining power is regional cooperation.

5.2 International Perspective in IIS Debate

To bridge the gap between the developed and developing countries' views on international internet connection market issues – specifically on the alleged inequitable and anticompetitive practices of top-tier IBPs - and to find solutions to the various problems we have indicated, Asia-Pacific ISPs and authorities will need firstly to come up with valid and concrete evidence to support their arguments, and resolve disagreements about analytic data-interpretation methods. These countries feel that developed nations are willfully delaying progress in these areas on the pretext that reliable data is lacking and more research or analysis is needed. However, Asia-Pacific countries will clearly be unable to gain broad public support from internet users, worldwide regulatory authorities and international organizations if they fail to make them appreciate the exact nature of their problems by providing unambiguous evidence about inequities in connection agreements. If internet interconnection problems are to be tackled through policy instruments or at legislative levels, regulatory bodies must first conduct investigation and analysis to determine the exact effect of these issues on the Internet market and end-users.

This in fact means that the international internet interconnection market as a whole must be analyzed, including internet traffic demand and supply, market structure, the status of peering and transit connection between suppliers (IBPs, etc.) and sources of demand (regional IBPs and ISPs, etc.). This analysis must be repeated periodically, and the results shared among interested parties. Moreover, to address specific questions highlighted in this chapter, the analysis must be broad-based and encompass information from both developing and developed countries: successful inter-regional cooperation will therefore be critical.

International organizations which are considered neutral and trustworthy, such as the ITU, APEC and OECD, are clearly the best to be in charge of this sort of regular survey and analysis. The first two of these are rather limited in terms of research capabilities in competition policy areas: in comparison, the OECD would be able to tackle the international internet interconnection issue in a more comprehensive fashion, across multiple levels of concern including telecom industry law, competition law and trade policy, and is therefore in a better position to conduct research and analysis on the interconnection market. Even though many Asia-Pacific countries are not members of the OECD, they are nevertheless allowed to participate in related research – but the cooperation of OECD member countries would be a *sine qua non*.

At the same time, Asia-Pacific organizations such as APEC and APT need to conduct research on internet interconnection market issues from their point of view, and report their findings regularly to international bodies such as the ITU, OECD and WTO. Results obtained from this effort can serve as a basis for Asia-Pacific countries to officially request that the EU and US also conduct studies on their side of the market[8]. This will require the creation of an inter-governmental council in charge of internet policy (an exiting or new sub-unit of APT is suggested) to oversee technical research relating to measuring internet traffic and conducting a structural analysis of the Asia-Pacific backbone connection market. APEC TEL (Telecommunications and Information Working

Group) has been carrying out research projects (APEC, 2005) related to measurement of internet traffic within Asia-Pacific region. However, developed countries have expressed reservations about the reliability of the group's results, and assigned them only a tangential importance as a simple reference. If they are to convey their views clearly and sell their arguments successfully to interested parties, Asia-Pacific countries must try to lobby to have such research to be carried across worldwide regions, and gain increased credibility for their views by entrusting project coordination to respected organizations such as the OECD and ITU.

Once tangible proof of anticompetitive IBP conduct or discriminatory/inequitable treatment of Asia-Pacific ISPs is obtained through international-level research and analysis of the internet backbone market, the next step is regulatory intervention to eradicate such practices at individual country level. The principal types of regulatory actions can be expected to be the designation of IBP facilities that qualify as 'essential facilities' and the introduction of preventive measures against anticompetitive conduct relating to connection, use or interconnection, or exclusionary practices by service providers. However, voices opposing such regulatory intervention can be heard in developing as well as developed countries, while implementing such competition-oriented regulation is often difficult in developing countries, and often cannot be achieved without technical and economic assistance from international organizations.

A working regulatory framework for stimulating competition is already in place in Korea. Although concrete models for pricing or settlement of connection fees are yet to be elaborated, directives intended to promote competition in the local internet market and protect the interest of local end-users have been already issued, including a requirement of disclosure of internet interconnection terms and conditions and pricing interconnection charge settlement guidelines etc. (MIC, 2005). Although the scope of these directives does not currently include the top-tier international ISPs, they are nevertheless expected to impact the international internet interconnection market, and these same principles should be applied to upstream market players as well.

6. CONCLUDING REMARKS

This chapter focuses on those current interconnection settlement arrangements that disadvantage Asia-Pacific ISPs and end-users. In order to understand the IIS, the hierarchy of internet interconnections and the various types of interconnection arrangements have been examined in detail. The main IIS issues raised in recent years between the Asia-Pacific and Western regions were summarized into 3 categories of concern; inequity in markets, anticompetitive practices and the prognosis of the balkanization of the Internet. In order to resolve these issues and to improve the IIS market in Asia-Pacific region, several recommendations have been discussed from both regional and international perspectives. As regional perspective actions, we have proposed 'continuous effort toward introducing greater competition into Asia-Pacific IIS market,' 'appropriate doses of regulatory intervention,' and 'active participation in international internet interconnection debates.' In relation to the last item we have provided a detailed strategy, and discussed related technical considerations. From the international perspective, we recommended two actions: 'collection and analysis of internet connection market information on a regular basis' and 'establishing rules and directives for promoting competition in the internet interconnection market.' In particular, we have emphasized that sharing structural information on the regional internet interconnection market between Asia-Pacific countries and enhancing the reliability of the market information analysis can give the entire region a more consequential voice with greater negotiating power in the international society.

The recommendations discussed by this study, based on its findings, presuppose close cooperation between concerned countries and market players, including regional ISPs. However, given that stakes are different from one country and one provider to another, even if similar problems are being faced, such cooperation may not easy to achieve. For instance, until the late 1990s, Japan agreed with other Asia-Pacific countries on most IIS issues. However, since the substantial growth of its internet backbone industry, it has veered away from its previous stance, and Japanese carriers' contributions in ITU-T SG3 meeting over the past several years (KDDI, 2003) have supported the position of the big North American and European IBPs. In another example, whilst the Australian regulatory authorities are among the most active advocates of the Asia-Pacific regional view, its largest provider (Telstra) now sides with big international IBPs (ITU-T SG3, 2003). In some ways these discrepancies are only to be expected: as the Internet evolves, so does its market landscape, and countries' and providers' strategies will change toward directions that reflect their own best interests. However, Korea has demonstrated in its recently-issued Information and Communication Public Announcement (MIC, 2005) that its government still espouses the Asia-Pacific view, even if the position of large Korean ISPs on international interconnection issues will probably come to differ significantly from the government position. Accordingly, the incorporation of this study's recommendations must be accompanied by a balanced consideration of both the future of the Internet and the interests of all sectors of the regional market, from IBPs to local end-users.

REFERENCES

AfrISPA (African Internet Service Providers Association) (2002). *The Halfway Proposition*, Retrieved Feb. 15, 2004, from http://www.afrispa.org/initiatives.htm

Antelope Consulting (2001). *The Costs of Internet Access in Developing Countries.*

APEC (2005). *Flow-based Internet Traffic Measurement and Analysis Project.* Final Project Evaluation Report, APEC #205-TC-01.5

Australia (2002). *Internet Traffic Exchange and the Development of End-to-End International Telecommunications Competition-Comments on the OECD Research Paper.* ITU-T Study Group 3, Delayed Contribution 22, COM 3-D 22-E, May.

Cable & Wireless (2000). Competition in the Supply of IP-network. A Position Paper by Cable and Wireless.

China (2002). *Proposed Changes to ITU-T Recommendation D.50 on New Settlement Principles and Systems for the Internet.* ITU-T Study Group 3, Contribution 16, COM 3-16-E, April.

CITEL (2001). *Response of Mexico to Base Questionnaire for Study of International Accords on Rates for Internet Service.*

Cade M. (2003). *Global International Internet Capacity and Traffic Data*. ITU-T SG3 Joint Rapporteur Groups On Internet TFM & IIC Meeting of Apr. 2003.

Cukier K. N. (1997). *Peering and Fearing: ISP interconnection and Regulatory Issues*, Retrieved Nov. 10, 2003, from http://ksgwww.harvard.edu/iip/iicompol/Papers/Cukier.html.

Elixmann D., & Scanlan, M. (2002). *The Economics of IP Networks – Market, Technical and Public Policy Issues Relating to Internet Traffic Exchange.* wik-Consult.

European Union, Commission (1998). Declaring a Concentration to be Compatible with the Common Market. *Official Journal of the European Commission.*

European Union, Commission (2002). *Contribution from the European Commission regarding ITU-T Recommendation D.50 and Associated Work in Study Group 3.* ITU-T SG3 document COM 3-D 24-E, May.

FCC (1997). *Digital Tornado: The Internet and Telecommunication Policy*. OPP Working Paper No. 29.

FCC (2000). *The Digital Handshake: Connecting Internet Backbones*. OPP Working Paper No. 32.

Intel (2000). *Caching for Improved Content Delivery*.

ITU-T (2000). *Telecommunication Reform: Interconnection Regulation*.

ITU-T SG3 (2003). *Report of The Joint Rapporteurs' Group Meeting Held in Brussels* on 20-21 Oct. 2003-Final version 5 Nov. 03-Rev.1," TU-T SG3 document, COM3-D53 Rev.2-E, Nov.

ITU-T (2004). *Asia-Pacific Telecommunication Indicators 2004*.

KDDI (2003). *A Suggestion for The Amendment of Recommendation D.50*. ITU-T SG3 document COM 3-D47-E, June.

Kende M. (2000). *The Digital Handshake: Connecting Internet Backbones.* Federal Communications Commission, Office of Plans and Policy, Working Paper no. 32, Washington DC. Sep.

Melian, J. P., Lopez da Silva, R. A., & Gutiererez, R. A. (2002). *Redes IP de nueva* generation.Communicatciones de Telefonica I+D, No.24

MIC (2005). *Interconnection Criteria of Telecommunications Facilities.* Public Announcement No. 2005-5 (in Korean).

OECD (2002). *Internet Traffic Exchange and the Development of End-to-End International Telecommunication Competition*. DSTI/ICCP/TISP(2001)5/Final, 13 March.

OECD (2006). *Internet Traffic Exchange: Market Developments and Measurement of Growth*. DSTI/ICCP/TISP(2005)11/Final, 05 April, retrieved Mar. 11, 2006, from http://www.oecdorg/dataoecd/25/54/36462170.pdf

OFTEL (2001). Effective Competition Review of Internet Connectivity.

Ovum (2002), IP Interconnect.

Probe Research (2003). *Global Internet Traffic-2003*.

RHK Inc (2003). *Asia-Pacific Capex to Decline, Traffic to Climb*. Annual Forecast Report, Asia-Pacific Telecom Dynamics.

Roseman D. (2003). *The Digital Divide and The Competitive Behavior of Internet Backbone Providers – A Way Forward.* Roseman Associates.

Sarrocco C. (2002). *Improving IP connectivity in the Least Developed Countries*, Retrieved May 17, 2004, from http://www.itu.int/osg/spu/ni/ipdc/index.html.

United States, General Accounting Office (US-GAO) (2001). *Telecommunications: Characteristics and Competitiveness of the Internet Backbone Market*, October.

World Trade Organization (1998), Telecommunication Services, Background Note by the Secretariat, S/C/W/74, 8 Dec.http://203.127.220.68/apecp1.nsf/websearch/D5038CDD0095E0348256BF400160256?OpenDocument

WEB SITES

http://www.afrispa.org/initiatives.htm

http://www.apec.org/apec/ministerial_statements/sectoral_ministerial/telecommunications/1998.html http://www.europa.eu.int/information_society/topics/telecoms/international/news/index_en.htm#editorial

http://www.itu.int/osg/spu/ni/ipdc/index.html

http://www.internetworldstats.com/stats.htm

KEY TERMS

Digital Divide: the gap between those people (or region, countries) with effective access to digital and information technology, and those (or region, countries) without access to it. It includes the imbalances in physical access to technology, as well as the imbalances in resources and skills needed to effectively participate.

Internet Balkanization: The deconstruction of universal internet connectivity. It can occur when a large number of small and medium-size ISPs are unable to connect to all parts of the internet

IIS (Internet Interconnection Service): A connection service between various Internet Service Providers including private and public service providers

Peering Agreement: An economic and technical agreement entered between two ISPs, on the exchange of routing information regarding traffic between internet networks.

Transit Agreement: A technical and commercial agreements that entered into between ISPs to provide transit services.

Transiting: Routing and transmitting of traffic by a transit service provider toward agreed-upon third parties.

ENDNOTES

1. See http://www.apec.org/apec/ministerial_statements/sectoral_ministerial/telecommunications/1998.html
2. See http://www.internetworldstats.com
3. Figure 1, 2 and 3 were made by using the data from “Asia-Pacific Telecommunication Indicators 2004” of ITU-T
4. See “Asia-Pacific Capex to Decline, Traffic to Climb,” Annual Forecast Report, Asia-Pacific Telecom Dynamics by RHK Inc (2003).
5. Figure 4 indicates only the international lease price. Transit services usually consist of lease line charges, port charges and other costs, e.g. bundle type. And the figure 4 was made by using the original data from TeleGeography, http://www.telegeography.com/ee/free’_resources/ and http://www.telegeography.com/ee/free_resources/essay-01.php.
6. The FCC currently does not require ISPs to report the content of peering or transit agreements: in other words, there are no rules regulating internet interconnection in the US (USGAO, 2001)
7. See http://www.wto.org/english/tratop_e/serv_e/telecom_e/tel23_e.htm
8. As of 2002, there had been no official requests for investigation of inequitable or anti-competitive behaviors in the international internet interconnection service market (Roseman, 2003).

Chapter X
A Techno-Economic Analysis of Mobile Virtual Network Operators

Dong Hee Shin
Penn State University, Berks Campus, USA

ABSTRACT

The purpose of this study is to examine cross-national data in order to identify possible factors related to the observable patterns of Mobile Virtual Network Operator (MVNO) penetration. For this purpose, a mobile market structure is analyzed considering important economic factors related to the penetration of MVNOs. MVNOs have gained popularity in the Western markets, but have a dismal track record in non-Western regions. In comparing the different regions, this study analyzes how the MVNO market has changed and what opportunities and/or threats network operators and potential MVNO entrants are likely to face. This study conducts an economic assessment of market structure and environment for different countries' MVNO penetration. The data on market structure is analyzed by means of cluster/factor analysis techniques in order to group countries according to their market environments. Then, regression equation analysis is used to investigate the relations of MVNO penetration and independent variables. The results show that MVNO penetration significantly relates to market structure and performance. These results also imply an appropriate policy to promote MVNO diffusion.

1. INTRODUCTION

Recent emerging telecommunication technology drives fundamental changes in the way the mobile industry does business. The ability to utilize emerging technology to offer a differentiated product or service experience to the mobile customer has become a key competitive advantage. Along with these benefits, technology has brought with it several challenges, including much more sophisticated and better informed customers, as well as the lowering traditional barriers to market entry. Most of all, emerging technology has a disruptive effect on mobile industry, which is best illustrated by Mobile Virtual Network Operators (MVNO): MVNOs are taking significant market share, changing supply chain of telecommunications, and are increasing churn and subscriber acquisition costs in mobile markets. Also, because many MVNOs have focused on offering discounted services, they have increased price competition and lowered ARPU (Average revenue per user).[1] A compelling question is how MVNOs have impacted the way markets are structured. Specifically, how MVNOs have altered existing business strategies and created new business models. These questions become more interesting with the trend that MVNOs have gained popularity in the Western markets, but have a dismal track record in non-Western regions.

Regarding this contrasting trend, when Virgin Mobile withdrew its operation from Singapore in early 2003, the company cited "weak economic and market conditions" in the country as the reason behind the failure of the business (Wireless Asia, 2002). Its decision to exit the Asian market was followed by the recent failures of MVNOs in Asian countries, such as Taiwan, Hong Kong, and Malaysia. As the Asian MVNO market development is not as visible as in Europe, the failures left some questions as to what market conditions would be required for MVNOs to succeed. Although there were other factors such as marketing strategy, prices, and culture that might contribute to the failure, this study focuses on market structure in a cross-national study to shed light on the question as to whether the observable differences between Eastern and Western market conditions are the outcome of a specific set of factors. This study investigates the set of factors influencing MVNO penetration in order to answer following questions:

RQ1: What are the market factors affecting the penetration of MVNOs?

RQ2: Is the penetration of MVNOs related to market structure?

When looking at the emerging MVNO phenomenon, there may be two approaches: (1) from the perspective of the MVNO itself, in which the features for its sustainable business development are researched; (2) from the perspective of the overall mobile market and industry, in which the relation of MVNOs to the market should be analyzed. The present study takes the second approach in order to analyze MVNOs from a broader perspective by relating MVNO penetration and market structure: (1) 28 countries of the EU and the 2 countries of the Americas representing the West, (2) 17 Asian countries representing the East, and (3) 21 countries of Latin and South America are collected. After collecting data on market and industry from these groups, cluster and factor analysis are used to summarize patterns of market structure in the mobile industry. Each factor is characterized by a set of coefficients expressing its correlations with the observed variables. The variables are assigned to the factor with which they are most loaded. As a result, the market structure indicators are split into disjoint sets, with each indicator associated with a single factor. The estimated factor loadings applied to the country-specific market structure indicators make it possible to score countries according to each of the factors, so that rankings of countries can be obtained in terms of factor-specific scores. It is standard practice to retain a number of the factors which cumulatively explain a substantial part of the overall covariance. A multivariate regression model is developed to evaluate the effects of market variables on MVNO penetration.

2. DEFINITION AND TREND OF MVNOS

There are wide ranges of definitions of MVNOs. The International Telecom Union (2001) defines an MVNO as a carrier providing users with mobile services without its own government-issued license for bandwidth. More specifically, Ulset (2002) defines it as an organization providing customers with mobile phone services without owning any airtime. This study takes the most recent MVNO definition from Shin and Bartolacci (2006): an organization that is more or less independent from its host network in terms of technology and service. It can be said that MVNO is a mobile service operator that does not have its own licensed spectrum and does not have the infrastructure to provide mobile service to its customers (i.e., it does not own the network on which its voice and data traffic is carried). Instead, MVNOs lease wireless capacity from pre-existing mobile service providers and establish their own brand names different from the providers. For example, Virgin MobileUSA uses Sprint as its underlying carrier, Page Plus uses Verizon, EZ Link Plus uses Cingular, and Air Voice Wireless uses AT&T.

The first commercially successful MVNO was Virgin Mobile UK, launched in 1999 and now has over 4 million customers in the UK. Its success was replicated in the US, but ventures in Australia have not been so successful, and failed in Singapore, albeit with a different pattern. There are currently approximately 200 operational MVNOs world-wide, with various models and concepts. Since 2000, MVNO has been evolved in two venues: one from original telecom operators such as Vodafone, and the other from retailer such as T-mobile and O2. MVNO from original telecom operators tend to provide cut-price call rates to market segments, whereas MVNO from retailers tend to focus on specific niche market demographics, such as Helio in the U.S. targets young users.

The number of MVNOs has been drastically increased since 2000. According to Ovum (2006), there were 59 MVNOs in Western Europe in 2002 and there are 190 MVNOs in 2006 in Western Europe. Countries like the Netherlands, France, Denmark, the UK, Finland, Belgium, Australia and the U.S. have the most MVNOs, whereas non-Western countries are just beginning to launch active MVNO business models. The MVNOs contained within their MVNO market study vary from consumer driven MVNOs to enterprise, non-consumer MNVOs, and data focused operations. The majority of MVNOs have been consumer focused and the future development of MVNOs will be within enterprise market developments.

Recently, MVNO providers, such as Virgin Mobile and ESPN, incorporate Wi-Fi to provide a wider breadth of services. Widening addressable markets and churn reduction are classic reasons for bundling new services, but practical reason for an MVNO to add Wi-Fi service is pent-up demand. Like wireline VoIP, a Wi-Fi phone integrated voice service could find a willing audience by lowering users' monthly bills. The Wi-Fi phone functions as a normal mobile handset unless in the range of a Wi-Fi hotspot and then it will use Voice over IP technology to make phone calls for a lower price. Thanks to Mobile Virtual Network Enablers (MVNE; an intermediary among host network operators, MVNO, and en-user customers), MVNOs can enlist help with the nuts and bolts of adding Wi-Fi if lacking a robust infrastructure.

The MVNO model has catalyzed vertical disintegration in the mobile industry and has created a situation where many major players separate from the mobile network operators, seeking to access the value chain at all levels (Shin & Bartolacci, 2006). In this light, Capuano (2005) likens MVNO to "liberalization frontier." Mobile network operators have benefits from MVNO by increasing network utilization and by using of excess capacity (Capuano, 2005). As these benefits of MVNOs are similar to the consequences of liberalization, privatizations, and deregulation, MVNOs can be located in this wider literature of telecom market and policy. Unlike other simple resellers of telecommunications services, MVNOs add value such as brand appeal, distribution channels, and other affinities to the resale of mobile services. In addition, unlike simple resellers, MVNOs have close and tight relationships with other players in the value chain. Effective MVNOs have sufficient agreements with existing operators to provide a good service coverage area, and some well-diversified MVNOs can offer a product mix that incumbent mobile operators cannot match.

Many studies have shown that the level of vertical integration is negatively related to competition and positively related to concentration (Hodendorn, 2005). The introduction of competition in the telecom market was followed by a decrease in market concentration, with new players entering into markets and a sharp drop in prices. Along with competition, liberalization has brought new players in both fixed and mobile markets. Market liberalization promotes competition by opening the market to new entrants backed by private capital. Liberalization is the opening of a monopolistic market to competitive provision of facilities and services (Melody, 1999). Therefore, liberalization and vertical integration are often negatively correlated (Shin & Bartolacci, 2006). As the new players develop new and niche services, the market becomes more and more segmentized. For most telecom players who are concerned with decreasing ARPU due to segmented markets and customers, the MVNO business is another way to increase ARPU from a stagnant customer base, with potentials and perils similar to broadband. MVNOs can increase ARPU by providing content-based ARPU and keeping their subscribers satisfied. As many studies show (Doyle & Smith, 1998; Noam, 2002), all these variables are closely related. With the emergence of MVNOs, it is worthwhile to examine each variable of market components in relation to MVNO penetration.

Although there have been many studies on MVNOs, few have attempted to look comparatively at MVNO effects and market structure. Most empirical studies on MVNOs are limited to single countries. This approach was partly related to the lack of internationally comparable data on market structures. A study by Boylaud and Nicoletti (2001) provides an example of how different market structures can be compared across countries. In addition, further studies (Noam, 2003a; Jang, 2003; Lehr, 2002; Dolye & Smith, 1998) helped to conceptualize the notion of market structure and link it to MVNOs. The present study extends the Shin and Bartolacci's (2007) study, which investigate MVNOs in four countries. By including virtually all cases of MVNOs in the world, the present study can locate MVNOs in a wider context, which would help demonstrate its unique contribution to the ongoing policy debates.

3. SUMMARIZING MARKET STRUCTURE FOR EMPIRICAL ANALYSIS

Data of a set of variables are drawn time period from 2004 to 2005. This period is the time when MVNO emerged and is rising to the present.

3.1. The Choice of Indicators

For the purpose of evaluating the correlations of market structure and MVNO penetration, it is useful to focus on a limited set of indicators that tom market structure/performance for 23 OECD countries. In the study, market structure was measured with the indicators of vertical integration, concentration, and segmentation: and market performance was measured with competition and liberalization. This study adds ARPU as a market performance measure, since the Federal Communications Commissions (FCC) commonly uses the ARPU measure when assessing mobile market performance.

In addition to these indicators, the pricing scheme, service demand, market maturity, governance mechanisms, and regulatory institutions can have important effects on performance.

However, these indicators are generally qualitative in nature and are difficult to convert to into quantitative values. In addition, such data are not sufficiently available over long enough time periods and across countries to be useful. The current regulatory environment associated with mobile telecommunications could be included, but it is characterized by a relatively low cross-country similarity regarding both price regulation and regulatory institutions. These dimensions of regulation and governance were therefore omitted. The focus thus was reduced to two main issues: market structure and industry performance. These variables were also used by the FCC (2005) when assessing mobile telecommunication markets; (1) market structure; (2) carrier conduct; (3) consumer behavior; and (4) market performance. Because of difficulty in collecting and turning the qualitative data on carrier conduct and consumer behavior into useful quantitative data, these variables are excluded. The following variables of the market structure were chosen:

- *Degree of liberalization.* The degree of liberalization is calculated as a percentage of the number of regulations, where no restrictions have been attached to both market access and national treatment for a specific services sub-sector and mode of supply out of total possible commitments. Calculations have been done by the GATT Trade Unit based on the methodology developed by Hoekman (1995). This methodology is in line with the method used in the telecom sector by Gual and Trillas (2004) who measure liberalization policies through the degree to which market opening or deregulation policies are asymmetric or biased in favor of entrants.

H1: The degree of liberalization of telecom market is related to the MVNOs' penetration.

- *Degree of vertical integration in the telecommunications industry,* as measured by the percentage of significant market players in mobile industries' revenue out of total telecom industry revenue. This measurement is used by Noam's (2004) study to investigate the Internet's vertical integration (percentage of top 10 companies in specific industries' revenue out of total industry revenue). According to Choi and Shin (1992), vertical integration illustrates how much an entity has created value on its own through its operating activities. Data on the degree of vertical integration were either (1) collected from various sources of ITU, OECD, and WTO, or (2) calculated from the Noam's (2004) method.

H2: The degree of vertical integration is related to the MVNOs' penetration

- *The degree of market segmentation,* as measured by the Market Segmentation Index (Shin & Bartolacci, 2006). The market segmentation method involves identifying and describing subgroups of the market. First, factor and cluster analysis profile customers. Based on survey responses, customers are grouped or segmented according to similar preferences or needs. Then, the Market Segmentation Index is computed by dividing the incidence of each market segment in a particular market by the incidence of each market segment in the total market. A score of 1.0 means that a market segment is the same as the total market in a market sample. Index values above 1.0 indicate that the market segment is over-represented in a given sample of the overall market, while index values below 1.0 indicate that the market segment is under-represented relative to the overall market in that sample. An index value of 1.2 means that market segment is 20 percent more likely to be found within the particular sample than in the general population.

H3: The degree of market segmentation is related to the MVNOs' penetration

- *The degree of mobile market performance*, as measured by average ARPU. Although there are other indicators of performance, such as CPGA (cost per gross acquisition), CCPU (cash cost per user), and churn, these data are hard to collect. Despite much criticism of its flaws, ARPU is widely used to measure market performance. Since the 1980s, the ITU has defined telecommunication performance indicators and ARPU has been a major indicator of performance (ITU, 2001). ARPU is calculated by dividing billable services-derived revenues for a given period by the average number of billable subscribers for that period.

H4: The degree of mobile market performance is related with the MVNOs' penetration

- *The degree of competition in the mobile telecommunications industry*, as compiled from data collected by the ITU and the World Bank (Standard Data) to construct a competition index variable. This variable is assigned a value of zero if the telecommunication sector is served by a national monopoly operator, a value of one if the sector has more than one operator in the mobile market segment, and a value of two if the sector has more than one operator in both fixed-line and mobile market segments. This method is used by Li and Xu (2002) for measuring competition in the telecom sector around the world.

H5: The degree of competition is related to the MVNOs' penetration.

- *The degree of concentration in the wireless telecommunications industry*, as measured by the HHI (Herfindahl-Hirschman Index), which is a commonly accepted measure of market concentration. It is calculated by squaring the market share of each firm competing in a market, and then summing the resulting numbers. The HHI number can range from close to zero to 10,000.

H6: The degree of concentration is related to the MVNOs' penetration.

3.2. Parametric Regression Analysis: Multivariate Regression Analysis

For estimation purposes, the reduced form of the conceptual model needed to be translated into an appropriate empirical model. Ideally, the model should be tested using time-series data, but the main constraints are related to the availability of data. As Bauer et al (2005) indicate, an international cross-section model raises other potential problems, and the existence of unique national situations may weaken the assumption that the same process links independent and dependent variables. This problem could be minimized using cross-sectional time-series data, which would allow estimating fixed-effect specifications that control for such factors. However, valid data on MVNO are only available recently. Thus, this study used a simple cross-sectional design based on observations for the year 2004 and 2005. The regression equation is derived from the aforementioned variables in a reduced form to explain MVNO penetration, as follows:

MVNO penetration = f (concentration, competition, vertical integration, liberalization, performance, and segmentation)

In this study, a multivariate ordinary least squares regression method was used to estimate model parameters. Factor analysis and cluster analysis methods were used in mapping the structure of the observations and relations between possible measures for the independent variables. The loading factors (0.6 or more) of variables are obtained from factor analysis. Based on the theoretical model mentioned before, the following empirical model was formulated:

MVNOP = *CONS* + β1**CON* + β2* *LIB* + β3* *ARPU* + β4* *SEG* + β5**COMP* + β*6VER*+ *ε*

with

MVNOP - MVNO penetration ; CONS - intercept (constant) ; CON - concentration (HHI); LIB – liberalization ; ARPU - Average revenue per unit; SEG – segmentation; COMP - competition; VER - vertical integration; ε – random error term.

The dependent variable was defined as the total MVNO market share in a country as a percentage of the total wireless market (the number of MVNO subscribers as percentages of the total mobile subscribers; Table 1e). Data were collected over a period between 2004 and 2005.

3.3. Cluster Analysis and Factor Analysis

Based on the aforementioned data, a cluster analysis was used to categorize and group mobile markets according to market structure and performance. Cluster analysis is used to investigate the similarity of the countries based on scores on each of the estimated factors. Starting from the values taken by the scores in each country, this technique progressively groups countries in clusters of increasing size, based on (multi-dimensional) pair-wise comparisons. At each step in the clustering process, an index of inter-group similarity measuring the distance between the clusters being joined is calculated. It is standard practice to stop the clustering procedure (i.e., select the relevant number of clusters) when the distance between clusters becomes sizable.

Using cluster analysis, several groups of markets were identified based on their experience with MVNOs over the period analyzed, the most vertical being Hong Kong and Singapore, and the most horizontal, layered or open (horizontally diversified or consolidated) being the UK, Denmark, Netherlands, Norway, Sweden, and the U.S. Country clusters based on market conditions were found to be as follows:

- Vertical structure (23 countries): Most Asian countries, Latin countries, etc.
- Horizontal structure (24 countries): Most of the EU, North America.
- Mixture (20 countries): Australia, New Zealand, etc.

After the groupings of countries were created through cluster analysis, a factor analysis was carried out to evaluate the underlying commitment factors and to reduce number of variables in order to avoid multi-colinearity (Table 1). The factors, which are linear combinations of the observed variables, can be interpreted using economic measures such as *market structure*, *openness*, and *dependency*. Each factor is characterized by a set of coefficients expressing its correlation with the observed variables, and the variables are assigned to the factor with which they are most loaded (Table 2). As a result, the market structure indicators are split into disjoint sets, each of which is associated with one factor. The estimated factor loadings applied to the country-specific market structure indicators make it possible to "score" countries according to each of the factors, so that rankings of countries can be obtained in terms of factor-specific scores. It is standard practice to retain a number of factors which cumulatively explain a substantial part of the overall covariance.

4. EMPIRICAL MODEL AND ANALYSIS RESULTS

4.1. Result from parametric analysis

Overall, the empirical results allow the identification of important factors that influence the

Table 1. Results of factor analysis: Rotated factor loadings

Variables/factors	Performance	Structure	Ancillary Regulation
Competition	0.628	0.052	-.0083
ARPU	0.721	-0.188	0.114
Segmentation	0.023	0.930	0.149
Concentration	-0.129	0.900	-0.033
Vertical integration	-0.093	-0.838	0.199
Liberalization	0.003	-0.254	0.349

**Factor loadings measure the correlation between the individual indicators and the latent factors. Indicators are assigned to the factor to which they are most correlated. The rotation of factor loadings is a transformation aimed at minimizing the number of indicators that are highly correlated with more than one factor.*

observable pattern of MVNO penetration. Four hypotheses were confirmed, while two hypotheses were rejected (H4 & H6). Four main factors are found to best describe the trans-national variance in the set of indicators of MVNO and market structure.

Estimation results summarize the estimated effects of market structure on MVNO penetration for the selected model specifications (i.e., those that were not rejected by the tests). Overall, the market structure indicators performed quite well, significantly improving the fit of the regressions. By looking at the indicators most closely associated with each of the factors, a straightforward economic interpretation can be made. The first factor (associated with vertical integration of the mobile market) shows that the degree of vertical integration is significantly associated with MVNO penetration; the presence of more vertically integrated incumbents reduces the MVNO market share. The second (associated with segmentation in telecom services) factor expresses the positive significant relationship with the MVNO penetration. The third factor (associated with mobile market competition) shows a positive relation, in that higher competition is related to higher

Table 2. Correlation matrix (full model)

	Concentration	Competition	Segmentation	Liberalization	ARPU	Vertical integration
Concentration	1					
Competition	-0.3923	1				
Segmentation	-0.2328	0.7992	1			
Liberalization	-0.2134	0.5923	0.4283	1		
ARPU	-0.3769	0.5912	0.6913	0.5131	1	
Vertical integration	0.5460	-0.4942	-0.4093	-0.7022	-0.4623	1

Table 3. Estimation results

Vertical integration	Segmentation	Competition	Liberalization	Concentration	ARPU	Constant	Prob>F R^2 Adj. R^2
0.061 *0.414*	**0.27442 *3.492*	0.0883 *1.349*	0.0302 *0.624*	0.0078 *0.029*	*0.0899 *3.392*	5.934 *1.170*	***0.0071 0.5517 0.4136

*Note: *significant at 90% level; ** significant at 95% level; *** significant at 99% level. (Dependent variable MVNO market share in 2002-2003, t-statistic is italic)*

MVNO penetration; and the fourth factor (associated with liberalization) accounts for the positive relation the process of the MVNO penetration. The concentration factor has a negative correlation with MVNO penetration, but the relation is not statistically significant. While ARPU does not show a statistically significant result either, ARPU yields an interesting insight. The ARPU of MVNO (breakdown by region) shows that North America has increased ARPU since 2000, indicating higher-end segments are coming online through an emphasis on post-paid MVNOs such as ESPN, Disney, and Helio. The ARPU of the EU MVNO has declined, implying growth in the low-end segments and intense competition. The ARPU of Asia is relatively stable, indicating a slight offset by data revenues from data-only operators.

Dividing the sample into the three groups identified by the cluster analysis yielded more insights (Table 4). The cluster analysis between the groups reveals that MVNO penetration among the groups correlates strongly with market structural factors (e.g., concentration, segmentation, liberalization, and vertical integration) and to a lesser extent with performance indicators (e.g., competition, ARPU).

The statistical analysis shows that the liberalization factor present in the horizontal group had a positive and significant impact on MVNO penetration. This can be explained by the EU's Access Directive, which helped to correct vertical integration in communications services and facilitated MVNO launches. In 2003, the European Commission issued a recommendation to national telecom regulators to examine the competitiveness of the market for wholesale access and call origination on public mobile telephone networks. The study resulted in new legislation in countries like Ireland and France that forces operators to open up their network to MVNOs. The EU's Access Directive contains requirements that vertically integrated network operators must impose accounting separation and initiate transfer pricing arrangements where they control access to essential input facilities that are used by their competitors in a retail market. In contrast, the liberalization factor for the vertical group was detrimental to MVNO penetration, although the effect also was not statistically significant. It can be inferred that the liberalized climate in Asian countries has provided established mobile operators with even more favorable markets. In other words, significant market players have been the biggest beneficiary of liberalization policy, while effective competition in the overall market has not been improved.

Model runs at the level of country sub-groups show that the overall pattern is particularly visible in the horizontal group of countries. In this group, competition is correlated positively with MVNO penetration (with no significance), and concentration does not exhibit a significant effect. Within the countries in the vertical group, competition has the opposite effect and is negatively related to MVNO penetration. Concentration (measured

Table 4. Estimation results, country groups

Variables/factors	EU	North America	Asia
Concentration	-0.08752 *-1.979*	*-0.01003 *0.9302*	*0.2101 *0.210*
Competition	*0.19832 *2.392*	0.1238 *1.341*	0.09333 *1.101*
Segmentation	*0.07832 *3.232*	**0.0932 *2.942*	*-0.0920
Liberalization	**0.2983 *4.344*	**0.3903 *3.9434*	**0.3683 *4.233*
ARPU	0.0738 *5.1323*	*0.0423 *2.381*	0.0128 *0.4923*
Vertical integration	**-0.1374 *4.999*	**-0.0393 *3.2832*	**-2.034 *4.213*
Constant	**12.3049 *4.923*	0.3943 *2.123*	0.984 *1.032*
Prob>F R^2 Adj. R^2	***0.0071 0.5517 0.4136	**0.0162 0.5711 0.4053	***0.0052 0.5291 0.4293

*Note: *significant at 90% level; ** significant at 95% level; *** significant at 99% level. (dependent variable MVNO penetration from 2004 to 2005, t-static in italics)* by HHI) shows a positive effect, but is not statistically significant on MVNO penetration. This finding contradicts a widely held belief exhibited in several previous studies (e.g., Noam, 2003b) that HHI is a standard indicator of concentration. The HHI of the U.S. is more effective than Europe. The HHI of the U.S. is 1,377 compared to that of Europe. Market concentration is low by international standards in the U.S., and the MVNO penetration, therefore, has been largely unaffected by market concentration. Concentration also is not a significant factor in Asia. For example, Hong Kong's HHI is low (1,936), but is an effective HHI figure. The HHI figures of the Asian countries contribute to the insignificant result of competition and concentration. As the competition variable shows, Hong Kong and Singapore are two of the world's most competitive mobile markets. The implication is that the Asian markets in general are small-scale markets, and therefore the incentive to innovate may be lower than in large markets. As shown in the EU and the U.S. mobile markets, larger firms innovate more. The largest operators in Hong Kong and Singapore have relatively small market shares. These cities host two of the most fragmented world markets, and less fragmented markets, such as the EU and the U.S., have much higher MVNO penetration. An explanation may be that fragmentized or segmented submarkets in smaller overall markets obviate inter-operator collaboration on application development (Godfrey & Kam, 2004). As to the HHI variable itself, it may not be a good measure in such a small market (Farrell & Shapiro, 1990; Demsetz, 1974). A more appropriate measure may be Gans' new measurement of HHI when it involves vertical structure, namely "vertical HHI" (Gans, 2005).

The market segmentation variable shows different results across groups. The segmentations in EU and North America show a significant correlation with the MVNO market share. This result reflects the current U.S.'s MVNE proliferation: there are more than 25 MVNE in the U.S. (Pyramid Research, 2006). An MVNE does not have a relationship with end-user customers. Instead, MVNEs provide infrastructure and services to enable MVNOs to offer services and have a relationship with end-user customers. MVNEs enable infrastructure and related services, ranging from network element provisioning, administration, operation support systems, and business support systems, by providing intermediation between MVNOs that do not want to have any control over network elements and those that want complete control. As of 2006, there are more than 39 MVNE's in the world; eighty percent of them are located in Western countries, and there is no MVNE in Asian countries (Pyramid Research, 2006). MVNEs accelerate market segmentation by allowing MVNOs to provide various segments such as prepaid, high-end, data-focused players, etc.

Interestingly, this segmentation variable shows a significant correlation with the vertical integration variable. This can be explained by the fact that the European MVNE's provide the telecom operators with a relationship to manage the ability to manage relationships through a single, trusted interface for different MVNOs (Kristensson, 2001). In other words, MVNE's provide a bundled platform for MVNOs. These equate to turnkey type solutions that typically include automated interfaces to a carrier's provisioning system, real-time or near-real-time usage, and gateways for Short Message Service and content, as well as the usual Telco back-office systems and services. MVNEs take such forms as Application Service Providers, billing companies, content providers, hardware manufacturers, and Internet or media companies. It can be said that MVNOs and MVNEs are motivated by the same factors but the latter influences the MVNO penetration (Shin & Bartolacci, 2006).

The highly vertically integrated markets in Asia have generally not been favorable to the MVNO model. The fact that the MVNO model has had difficulty establishing itself in such segmented market environments suggests that the intensity of the facilities-based competition creates limited opportunities for them to be economically viable, even with differentiated strategies in the face of aggressive price competition.

In regard to market performance, the ARPU variable shows no significant correlation, although it does show a positive correlation. The insignificance of ARPU implies that the main challenge prospective MVNOs will face is not necessarily a low ARPU, but other challenges such as a niche market. In other words, regardless of whether mobile operators are making or losing profit, they are seeking new markets. As Jost (2004) predicts, MVNOs can potentially survive on $5 ARPU, but the MVNO model in general must find a way to survive on the cost side of the business. The positive relation implies that MVNOs are active in markets where they deal with low margin subscribers. The undesirable role of MVNOs in this scenario protects the host carrier from undesirable financial costs such as those related to volatile prepaid churn, high capital investment, and consumer fraud. All of these can accompany an MVNO that targets low-value subscribers. The positive correlation further suggests that MVNOs must find a niche among incumbent carriers who are willing to lease capacity, since most operators are unlikely to open their networks.

The estimation results of liberalization show that liberalization is a somewhat significant factor (Table 3). It is interesting to note that the Asian markets' liberalization also significantly contributes to MVNO penetration (Table 4), despite the low penetration in general. Since the market liberalization in the EU, a huge number of competing MVNOs are operating, while in Asian markets, under pressure of regulations and entrepreneurs, the process of MVNO development has just started. This can be interpreted that several Asian countries like Japan, Malaysia, and India are making positive efforts to facilitate MVNOs by establishing guidelines requiring network operators to lease or remain open to MVNOs.

The contradiction between liberalization and low MVNO penetration in Asian countries also can be explained from the different nature of liberalization. Asian countries in general achieved liberalization in policy and regulation, but actual markets seem not to fully feel such liberalized environment (Chowdary, 1998). While the liberalizations in Western countries have been achieved in an autonomous interaction between government and market, Asian countries seem to achieve it solely by policy-driven liberalization (Shin, 2007). The liberalization indicators in Asian countries may be policy-based, which is different from the market-based liberalization of Western countries. Under the policy-based liberalization, the markets are artificially changing toward liberal markets, but liberalized markets are most effective in climates of effective competition (Wu, 2003). Although overall Asian markets are apparently liberalized by government intervention, Asian markets still seem to be not liberal enough to embrace the MVNO, and the MVNO does not appear to stabilize in markets without effective competition.

Overall, the empirical analysis does show key drivers of MVNO penetration, but it also implies that several constellations of factors can be conducive to MVNO penetration. The active operation of MVNEs supports the analyses of segmentization, vertical integration, concentration, and liberalization. This point needs to be further examined by using panel regression.

5. CONCLUSION

The emergence and proliferation of MVNO is not simply matter of infrastructure sharing between HNO and MVNO. It may involve a more fundamental question of market condition as the emergence of MVNOs has revealed two main industry structures in the mobile telecommunications industry: a horizontally layered (either diversified or consolidated), market-driven structure with a modular product architecture, and a vertically integrated, walled-garden oligopoly with an integrated product architecture. As shown previously, MVNO services are much more successful in Western than in Asian countries. One major reason for this difference in success is the differences in the industry structures between the markets in both areas. Because of the complexity of MVNOs, a horizontally layered structure seems to be more inviting. This finding provides future carriers and potential MVNOs with a framework to determine whether or not the MVNO model is right for their respective business.

MVNO services are more successful in markets where virtual operators take intermediary and aggregating roles in order to offer a true end-to-end mobile service. The incumbent Asian operators are not eager to change the market structure. It seems that MVNOs would reap nothing but negative consequences from the consumer's point of view if the market structure changed. In his study on market structure, Economides (1996) leaves open the question of whether the horizontal or a vertical industry structure is better for MVNOs. That is, the open questions would be: Is it more valuable to have a transparent, market-driven industry structure or a vertically integrated oligapoliy enabling the provision of easy-to-use services? Would this more integrated structure provide a mobile telecommunications market with only one, monopolistic player. This study partly answers such questions by showing that a horizontally-layered market is better at facilitating MVNO penetration. It appears that the vertically integrated telecom business models are less effective in an environment where the demand for a whole range of value-added services is strong. Under such a horizontal value chain, vertical integration seems to be minimized. Predicting which MVNO trends will prevail can be a hard question. In the long run, MVNOs, along with MVNEs, are likely to rise and decline in popularity as uncertain markets and industries and change. One thing is clear that regardless of whether MVNOs, prevail or fail, their impact will be significant throughout the wireless industry.

6. LIMITATIONS AND FUTURE STUDIES

Given its complexity, defining telecom market structure is always elusive and difficult. This study is an exploratory attempt to define mobile market structure to analyze the emerging phenomena of the MVNO. For parsimony, the study neglected various issues like the differences between the service demand scenarios in different countries and the maturity of the market. In addition, this study used an aggregation model, and data were collected for the limited time of 2004-2005. Therefore, it would be helpful in the future to conduct comparative studies at a less aggregated level, such as using cross-sectional and time-series analysis methods. In addition, multivariate regression, as Cava-Ferreruela et al. (2006) indicate, has weak numeric and estimation features. Other sophisticated measures may be accompanied with multivariate regression, such as bootstrap re-sampling or stepwise selection techniques.

Furthermore, it is not desirable to set a linear relation between the dependent variable (MVNO penetration) and independent variables in this study, because the relation may not linearity. The relation, however, can be natural log (non-linearity). Using natural log of non-linearity, the analysis will be smoother and produce better ideas. Future studies may complement the nonparametric local polynomial regression techniques to the existing parametric regression model. Although the relation may not be quite linear, the existing parametric regression analysis is worthwhile to keep for comparison to the non-parametric, given that the relation can be largely used in other studies in telecommunications. Using both parametric and nonparametric regression, theoretical robustness can be increased. In addition, a strong conceptual framework will better predict the prospect for MVNOs.

The current study intentionally simplifies and limits the variables in order to grasp an overall understanding of MVNO penetration factors. The limited variables allowed this study to set aside regulations, and policy, and carrier conduct and behavior as dummy variables. As Economides (1996) argues, the structural differences are caused by dissimilarities in the national regulations. As the present study excludes national policy from variables, it neglects to note many important aspects: EU regulators are opening up network access in favor of MVNO models and the FCC has been generally favorable to MVNO, which is not the case in non-Western regions. In addition, the significant results of the dummy variable imply that there is something more behind MVNO penetration and market structure. For example, despite the overall similarities between the U.S. and the EU, the lower MVNO in the U.S. can be explained by the fact that the U.S. operators are carefully watching MVNO development because they believe that the transition period between 3G and 4G will be the most fruitful period for MVNOs. This kind of carrier behavior will have to be investigated with respect to regulation and policy.

REFERENCES

Bauer, J., Kim, J., & Wildman, S. (2005). *Effects of national policy on the diffusion of broadband in OECD countries*. The UFL-LBS Workshop of the Future of Broadband: Wired and Wireless. Gainesville, FL, February 24-25, 2004.

Boylaud, O, & Nicoletti, G. (2001). Regulation, market structure and performance in telecommunications. *OECD Economic Studies*, *32*, 99-142.

Capuano, C. (2005). Access regulation in vertically integrated oligopolies. *Proceedings of 32th the European Association for Research in Industrial Economics (EARIE) Conference*.

Cava-Ferreruela, I., & Alabau-Munoz, A. (2006). Broadband policy assessment: A cross-national empirical analysis. *Telecommunications Policy*, article In Press.

Choi, C., & Shin, H. (1992). A comment on a model of vertical product differentiation. *The Journal of Industrial Economics*, *40* (2), 229-231.

Doyle, C., & Smith, J. C. (1998). Market structure in mobile telecoms: qualified indirect access and the receiver pays principle. *Information Economics Policy*, 10(4), 471-488.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization*, *14*(2), 675-699.

Farrell, J., & Shapiro, C. (1990). Horizontal Mergers: An Equilibrium Analysis. *American Economic Review*, *80*(1), 107-123.

Gans, J. S. (2005). Concentration-based merger tests and vertical market structure. *Journal of Economic Literature*, *43*(2), 238-256.

Godfrey, J., & Kam, J. (2004). *What it takes to kick start mobile data: the lessons from Japan and South Korea*. ITU Asia 2004, Online News Service.

Gual, J., & Trillas, J. F. (2004). *Telecommunications policies: determinants and impact*. (August 2004). CEPR Discussion Paper No. 4578. Available at Social Science Research Network: http://ssrn.com/abstract=604468

Hodendorn, C. (2005). Regulating vertical integration in broadband: Open access versus common carriage. *Review of Network Economics*, *4*(1), 19-32.

Hoekman, B. (1995). *Trade laws and institutions: Good practices and the World Trade Organization*. Supplementary paper presented at The Uruguay Round and Developing Economies, *a World Bank Conference*, 26-27 January 1995.

Informa Research (2005). *Informa research services' telecom market segmentation and customer profiling program*. Accessed in March 2006 http://www.markettrends.com/industry/telecom/telecomproj.htm#5

International Telecom Union (2001). *Mobile Virtual Network Operators.* Geneva. Geneva: ITU.

International Telecom Union (2004*). Trends in telecommunication reform: Effective regulation*. Geneva: ITU.

International Telecom Union (2002). *Asia-Pacific Telecommunications Indicators.* Geneva: ITU.

Jang, B. (2003). Vertical structure and network competition in telecommunications services market. *Information Society Research,* Fall: 81-101.

Jost, J. (2004). *Can MVNOs succeed in emerging markets?* Pyramid Research.

Krafft, J. (2003). Vertical structure of the industry and competition: an analysis of the evolution of the info-communications industry. *Telecommunications Policy*, *27*(8/9), 625-649.

Kristensson, S. (2001). *Mobile Virtual Network Operators. Assessing MVNO business opportunities.* School of Business, University of Stockholm, Master of Science Thesis, June 2001.

Lehr, W. (2002). Vertical integration, Industry Structure, and Internet Telephony. In McKnight, Lehr, and Clark (Ed.), *Internet Telephony*. The MIT Press.

Li, W., & Xu, L. C. (2002). The political economy of privatization and competition: Cross-country evidence from the telecommunications sector. *Journal of Comparative Economics*, *30*(3), 439-462.

Melody, W. (1999). Telecom reform. *Telecommunications Policy, 25*, 7 – 34.

Noam, E. (2004). The Internet: Still Wide Open and Competitive? Oxford Internet Institute, Internet Issue Brief No. 1, August 2003.

Noam, E. (2003). The effect of deregulation on market concentration: an analysis of the Telecom Act of 1996 and the industry meltdown. *The Columbia Science and Technology Law Review*, *4*, 1-12.

Noam, E. (2003b). Deregulation and market concentration: An analysis of post-1996 consolidations. *Federal Communications Law Journal*, *58*, 539-550.

Noam, E. (2002). *How telecom is becoming a cyclical industry, and what to do about it*. Presentation at Columbia Institute for Tele-Information, Conference on Telecommunications Volatility, April 2002, Columbia University, New York.

OECD Economics Department Working Papers (2001). OECD Economics Department in its series by Olivier Boylaud and Giuseppe Nicoletti. Regulation, market structure and performance in telecommunications, 2001.

OECD (2004). Performance Indicators for Public Telecommunication Operators. Paris. Original OECD report available online at: http://www.oecd.org/dsti/sti/it/cm/prod/ICCP22.htm

Ovum (2006). Virtual Mobile Services: Strategies for Fixed and Mobile Operators. Ovum Ltd. 2006, by John Matthews, Susan Sweet. http://www.ovum.com.

Pyramid Research (2006). *MVNOs and MVNEs: Analyzing the Viability of Virtual Mobile Players.* February 2006 (ID: PYR1224156) Accessed in August 2006. http://www.marketresearch.com/product/display.asp?productid=1224156

Sappington, D., & Weisman, D. (1996). *Designing incentive regulation for the telecommunications industry*. MIT Press, Cambridge, MA.

Shin, D. H., & Bartolacci, M. (2007). A study of MVNO diffusion and market structure in the EU, US, Hong Kong, and Singapore. *Telematics & Informatics, 24,* 86-100.

Shin, D. (2007). A critique of Korean national information strategy: case of national information infrastructures. *Government Information Quarterly, 24*(3), 624–645.

Ulset, S. (2002). Mobile virtual network operators: a strategic transaction cost analysis of preliminary experiences. *Telecommunications Policy,* 26(9-10), 537-549.

Wireless Asia (2002). Virgin Mobile and SingTel jointly call off their Virgin-branded MVNO mobile service a mere eight months after its launch. August 21 2002.

Wu, F. (2003). From monopoly to competition: a comparative study of telecommunications policy reform in four Asia countries. *Proceeding of Pacific Telecommunications Council,* January 2003, Manoa,Hawaii.

KEY TERMS

ARPU (Average Revenue Per User or Average Revenue Per Unit): This term is used by telephone carriers. This term has also become widely used in other industries including consumer Internet service providers. It is the revenue generated by a customer phone, pager, etc., per month.

Herfindahl-Hirschman Index (HHI): A commonly accepted measure of market concentration. It is calculated by squaring the market share of each firm competing in a market, and then summing the resulting numbers. The HHI number can range from close to zero to 10,000.

Horizontal Markets: Two markets are horizontally related if the two goods in question are substitutes or complements in production or consumption.

MVNO (Mobile Virtual Network Operator): A company that provides mobile phone service but does not have its own frequency allocation of the radio spectrum, nor does it have all of the infrastructure required to provide mobile telephone service.[I]

MVNE (Mobile Virtual Network Enabler): A MVNE does not have a relationship with end-user customers. Instead, a MVNE provides infrastructure and services to enable MVNO's to offer services and have a relationship with end-user customers.

Vertical Integration: A style of ownership and control. The degree to which a firm owns its upstream suppliers and its downstream buyers determines how vertically integrated it is.

Vertical Market: A vertical market is a particular industry or group of enterprises in which similar products or services are developed and marketed using similar methods (and to whom goods and services can be sold).

ENDNOTE

1 ARPU is the revenue generated by a customer phone, pager, etc., per month. In mobile telephony, ARPU includes not only the revenues billed to the customer each month for usage, but also the revenue generated from incoming calls, payable within the regulatory interconnection regime.

Chapter XI
Evolving Value Networks and Internationalisation of National Telecommuncation Companies from Small and Open Economies

Riku Laanti
The University of Adelaide Business School, Australia

Fred McDougall
The University of Adelaide, Australia

Georges Baume
The University of Adelaide Business School, Australia

ABSTRACT

This chapter focuses on the internationalisation processes of national telecommunications companies (telcos) from small and open economies (SMOPECs) who have moved from a domestic monopoly to an actor within the global industry. This chapter aims to increase our understanding of how these companies have internationalised, what factors have been the most influential in this process, and how the position of these companies has changed in the evolving value network of the industry. The basis of the analysis will be a theoretical discussion about the concept of value networks and research on the internationalization process of a firm. The study reported in this chapter is part of a wider research project on the internationalisation strategies of telcos from SMOPECs. Case examples from that study will be used to illustrate the internationalisation processes of telcos from SMOPECs within the context of the whole industry value network. Finally, conclusions will be drawn and future research opportunities suggested.

INTRODUCTION

The telecommunications industry has experienced significant transformation: value chains have changed from vertical to horizontal, and there has been an important shift from value chains to value networks (Cave & Waverman, 1999; Fjeldstad, Becerra, & Narayanan, 2004; Li & Whalley, 2002; Sabat, 2002; Steinbock, 2003). Many factors, such as digitalisation, deregulation and privatisation, have contributed to these developments (Häikiö, 2001; Ramamurti, 2000). Li and Whalley (2002), who explored the complexity of value networks, argued that new research is needed in this still evolving area. They emphasized that ”To survive and thrive in this new environment, every company needs to understand their positions in each of the value chains within the value network, and to re-evaluate their strategies and business models” (Li & Whalley, 2002, p. 469).

This chapter will analyse these developments by focusing on the internationalisaton processes of national telecommunications companies (telcos) from small and open economies (SMOPECs), who have moved from a domestic monopoly to be an actor within the global industry. SMOPECs include countries such as Austria, Denmark, Finland, Ireland, New Zealand, Portugal, Norway, Sweden, and Switzerland, who have integrated in the world economy by lowering or eliminating their trade barriers (Benito, Larimo, Narula, & Pedersen, 2002; Kirpalani & Luostarinen, 1999; Maitland & Nicholas, 2002; Merrett, 2002). Multinational companies (MNCs) from smaller countries face specific challenges due to their relatively smaller size and limited resources, especially in capital-intensive service industries such as in the telecommunications industry. In these types of industries where expensive foreign direct investments (FDIs) are commonly the mode to enter international markets, the largest MNCs from the largest economies of the world often dominate the sector (Knight, 1999). Research findings are still limited on how MNCs from smaller countries have managed these challenges and if there are alternative strategies available for them internationally (Knight, 1999). This chapter aims to increase our understanding of *how* these companies have internationalised, *what factors* have been the most influential in this process, and *how the position* of these companies *has changed* in the evolving value network of the industry. An analysis of telcos from SMOPECs provides valuable longitudinal empirical data on this topic, as many of these companies were among the first telcos to internationalise.

The study described in this chapter is part of a longer ongoing research project on the internationalisation of telcos from SMOPECs (Laanti, 2008). The purpose of the project has been to analyse the applicability of traditional internationalisation theories, especially internationalisation process theories and the latest strategic theories on globalisation, to a service network industry. Case examples from the empirical data of the wider study will be used here to illustrate the different phases of the internationalisation process of telcos from SMOPECs within the context of the whole industry value network.

The findings described here should be useful to researchers and managers of internationalising telcos from SMOPECs, and telcos more generally, for other companies in the telecommunications industry or the whole ICT industry's value network, and for policy makers to increase their understanding of the internationalisation process and associated factors.

The next section focuses on the concepts of value chain and value network, and discusses the telecommunications industry´s value network and telco's position and role in it. It is followed by a brief review of the research on internationalisation theories, with a more specific description and analysis of the internationalisation process of telcos from SMOPECs. Finally, conclusions are drawn and future research opportunities suggested.

FROM VALUE CHAINS TO VALUE NETWORKS

Research on Value Networks

The concept of *value chain* has been widely covered in the strategic management literature, most notably by Porter (1985). A value chain arises from a company's activities and internal business processes to create value for its customers (Porter, 1985). This company specific value chain also belongs to a broader industry value chain or system (Porter, 1985). A company must recognise its most important value chain activities, how to add most value for its customers as well as how to manage its more generic value chain activities. To do this, managers must understand both the external environment, including industry value systems and customers, and the key internal resources and core competences. However, although these first studies on value chain recognised the importance of the external environment of a firm, they did not have their focus on the international aspect and impact of value systems.

Recent developments in the telecommunications industry, not least due to globalisation developments, have seen value chains moving from vertical to horizontal, and from value chains to value networks (Li & Whalley, 2002; Sabat, 2002). The difference in the definition of value networks, when compared to the traditional definition of value chains, is that in a value network there are several entry and exit points, and that activities take place simultaneously, instead of successively (Li & Whalley, 2002). Fjeldstad et al. (2004) argued that value network models are especially successful in describing the business operations of companies such as telcos, which use mediating technologies.

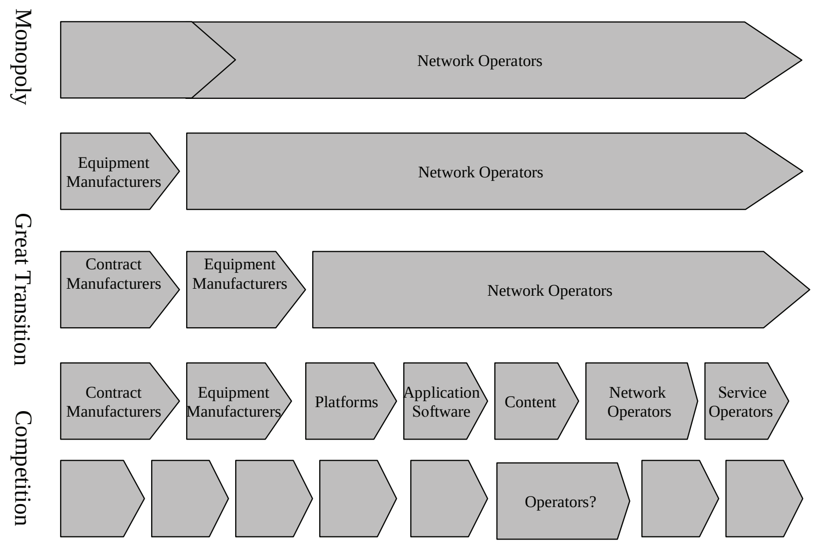
The Telecommunications Industry's Value Networks within a Single Country

Two notable developments in the telecommunications industry have occurred. First, value chains started to disintegrate from a monopoly situation, in which they were highly vertically integrated (Contractor & Lorange, 2002; Economides, 1996; Glachant, 2002). Historically, national telecommunications industries were dominated by monopoly telcos, which basically operated the whole value chain from R&D, to equipment manufacturing, to network operating, to service operations, to the sales of the telecommunications equipment to the end-customers. A well documentend example is AT&T in the US (Fransman, 2002).[1] However, deregulation developments gradually opened competition and monopolistic value chains started to disintegrate including more independent players. Competition increased in the equipment manufacturing/supply side to telcos and in the sales of equipment to the end-customers, and finally in the actual telco operations as well. This resulted in shifts up and down the value chain.

Second, convergence across the telecommunications, IT, and media industries has caused value chains to integrate horizontally (Sabat, 2002), resulting in a new and broader definition for the whole industry - the information and communications technology (ICT) industry. Also, new significant services were developed, notably Internet and mobile communications. All these developments changed radically the operating environment of traditionally monopoly-based telcos. They faced challenges in the form of competition and at the same time new opportunities arose in new business areas.

While existing players repositioned to new and more focused business areas, new entrants also emerged in the ICT ecosystem (Sabat, 2002). These included software developers, new infrastructure suppliers, and content and service

Figure 1. Vertical integration of value chains into value systems (adapted from Steinbock (2003))



providers, such as banking by mobile and over Internet systems, mobile and online games, and videos and other information services (Peppard & Rylander, 2006). Opportunities emerged for the players who were able to identify and utilise the 'sources of value in the network' (Peppard & Rylander, 2006, p. 139).

To demonstrate this progress in the wireless technology industry, Steinbock (2003) defined three different development phases: monopoly, great transformation, and competition (see Figure 1). In the *monopoly phase*, the whole value chain was operated by one company, a national monopoly telco. During the *great transformation*, though, independent equipment manufacturers started to take a more important role[2]. In the *competition phase*, the value chain disintegrated even more and the relative importance of the national telcos decreased significantly. Similar developments could be seen in other business areas in the industry, such as data communications.

This new competition brought challenges to ex-national monopoly telcos, as their share of the total revenues and market shares of each individual part of the business started to shrink considerably. They were able to compensate this at some level by entering new rapidly growing business areas such as mobile and Internet. However, they were also forced to start looking for new growth opportunities across their national borders.

INTERNATIONALISATION DEVELOPMENTS

As aforementioned, increasing competition in the domestic markets acted as a push force for telcos to look for growth opportunities internationally. Moreover, globalisation developments opened foreign markets and made it easier for companies to internationalise. In the 1980s and 1990s globalisation seemed to be a fad in management research and practice, which may have put further pressures on telcos to internationalise.

Research on internatonalisation has been covered in both the strategy and international business literature. Moreover, in international business two separate research streams have studied firms' internationalisation, namely the economic and process streams (Benito & Welch, 1994; Liesch et al., 2002). In this chapter, the focus is on process theories, although findings are also drawn from the economic stream and from the strategy literature.

Most traditional theories on the internationalisation process of a firm (Johanson & Vahlne, 1977; Johanson & Wiedersheim-Paul, 1975; Luostarinen, 1979, 1994) are based on research conducted on manufacturing MNCs. These process theories suggest that firms start their internationalisation by entering foreign markets with small psychic distance; that is, cultural and geographically close countries (Johanson & Vahlne, 1977; Johanson & Wiedersheim-Paul, 1975). Also firms start their internationalisation with less-committed operation modes, such as export, before committing more resources by investing in foreign markets. The process is linear and eventually develops into a global strategy by a MNC (Gabrielsson & Gabrielsson, 2004).

In the 1980s and especially in the 1990s, globalisation was one of the major topics of research in management studies. Several researchers argued that companies need to go global as consumer tastes and markets converge rapidly (Levitt, 1983; Ohmae, 1989). Yip (1989; 2000), in his seminal works on globalisation strategies of companies, classified globalisation drivers into four major groups: cost, market, competitive and governmental drivers. Within manufacturing sectors some of the key drivers identified were cost drivers such as economies of scale in R&D and production, and sourcing materials; and market drivers such as searching for new customers in foreign markets. For example, to balance escalating R&D costs, manufacturing companies need to achieve greater sales volumes, thus needing to deliver products outside of domestic markets. This need to internationalise seems to be most evident for manufacturers from smaller countries with limited domestic demand (Steinbock, 2003).

Another way to analyse the internationalisation process is to focus on organisation strategies. Bartlett and Ghosal (1992) argued that firms' strategies develop from domestic, to international, to multinational, to global, and finally to transnational. International strategies relate to companies which mostly export their products overseas without investing in significant foreign operations. Multinational strategy treats each country organisation individually, whereas in a global strategy operations are strandardised and centrally managed. Transnational strategies try to combine the ability to adapt locally, but to draw resources and share knowledge globally.

Telecom equipment manufacturing companies seem to fit nicely into these theories, as they first started to internationalise incrementally, often by exporting their products, but soon established production facilities internationally. Also, these large MNCs seemed to implement global strategies relatively rapidly (Steinbock, 2003). However, there are several challenges to these traditional theories, such as the internationalisation of born global companies and service internationalisation. Even within the same industry, these very different types of companies have followed different strategies in their internationalisation.

Born Global Companies

A group of companies that have challenged many traditional internationalisation theories are born globals. They are often small companies which, despite limited resources, are able to internationalise very rapidly, often to countries with large psychic distance (Knight & Cavusgil, 1996; McDougall, Shane, & Oviatt, 1994; Rennie, 1993). To manage this rapid internationalisation born globals often follow international/global niche strategies. Also, in many cases digitalisation and the use of Internet has been a major factor enabling this rapid internationalisation process for these companies (Laanti, Gabrielsson, & Gabrielsson, 2007). In the ICT sector many small but rapidly internationalised application developers and content providers could be labelled as 'born globals'.

Service Internationalisation

Another challenge to traditional internationalisation theories that has attracted debate among researchers has been the internationalisation of services. Several researchers have argued that there are a number of significant differences in how services and manufacturing companies internationalise (e.g. Brouthers & Brouthers, 2003; Contractor & Kundu, 2000; Erramilli & Rao, 1993). First, service companies started to internationalise much later than most manufacturing companies. This was partly due to the different treatment of services in the world economy, as it was not until the Uruguay round of GATT negotiations in 1993, and the subsequent lowering of trade barriers for services, that resulted in a growth in services trade (Clark, Rajaratnam, & Smith, 1996; Javalgi, Griffith, & White, 2003).

Second, many services have followed their industrial clients abroad, resulting in *follow-the-customer* strategies, with more rapid processes than traditional theories would suggest (Majkgård & Sharma, 1998; Roberts, 1998).

Third, services also possess several unique characteristics, which may cause them to be more challenging to internationalise than goods. Researchers have identified characteristics such as *intangibility* (Boddewyn, Halbrich, & Perry, 1986; Clark et al., 1996; Erramilli, 1990; Javalgi et al., 2003), *inseparability* (Erramilli, 1990; Javalgi et al., 2003), *heterogeneity* (Erramilli, 1990; Knight, 1999), and *perishability* (Erramilli, 1990; Lovelock & Yip, 1996). Thus, given the previously mentioned, 'exporting' as an operation mode may be infeasible in many service sectors (Erramilli, 1990; Javalgi et al., 2003).

Four, services are also very heterogeneous, which means that there are also differences across different service sectors in regard to their internationalisation strategies. Thus, several researchers have also classified services based around their business processes and service characteristics, such as Erramilli's (1990) *hard- and soft-services;* Boddewyn et al.´s (1986) three categories of *foreign-tradable services, location-bound services*, and *combination services*; and Clark et al.'s (1996) *contact-based services, vehicle-based services, asset-based services*, and *object-based services.* For example, using Erramilli's classification a music-CD is a hard-service and is thus more easily exportable than a concert which is a soft-service.

Telco Internationalisation

Most services that telcos provide are similar to many other network industries, such as utilities and railways; that is, a physical network is required before the service can be provided (Economides, 1996; Shy, 2002). Thus, it can be argued that these services are mostly *soft-services, asset-based and*/or *location-bound services.* Moreover, in the telecommunications service industry there were also several characteristics typical of these and other network industries such as network externalities and the strong role of government (Economides, 1996; Liebowitz, 2002; Shapiro & Varian, 1999; Shy, 2002). Also Yip (1989), in his research on globalisation, emphasized that in services the role of governmental drivers is much greater than in manufacturing industries.

All this means that the logic in the internationalisation strategies for telcos may vary significantly from that for manufacturing companies (Sarkar, Cavusgil, & Aulakh, 1999), or for a service which is more closely related to goods, such as a hard-service. Following this reasoning, telcos have not been able to follow similar internationalisation processes to telecommunications manufacturers or to that of many 'born globals'.

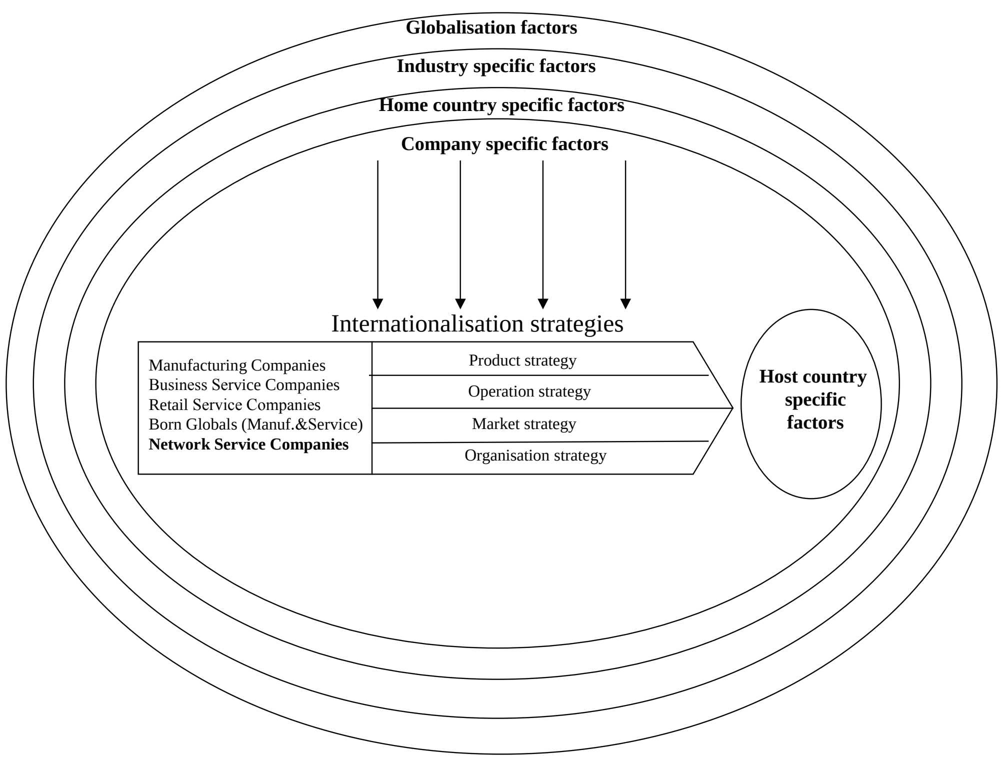
This also supports the findings of several other researchers (see e.g. Lovelock & Yip, 1996; Majkgård & Sharma, 1998), who argued that different industries follow different internationalisation paths, and research on internationalisation should pay more attention to idiosyncracies and context specifity. In this context, Laanti et al.

(2006) developed a framework to analyse internationalisation strategies of telcos (see Figure 2). In this framework, five major groups of factors that influence internationalisation strategies were identified: globalisation factors, industry specific factors, home country specific factors, company specific factors, and host country specific factors. It was argued that there may be significant differences in how companies internationalise, based on differences in these factors. The focus of this chapter is on industry specific factors, and also on home country specific factors. The study that forms the basis of this chapter is focused on telcos from SMOPECs, and the influence of these factors on the internationalisation strategies of telcos.

Research with a focus on the internationalisation process of telcos is still relatively sparse. Some important studies were undertaken by Sarkar et al. (1999) and Stienstra et al. (2004). Sarkar et al. argued that telcos' internationalisation process is very unique due to three factors: the industry's oligopolistic structure, network characteristics, and limited opportunities available in foreign markets. This results in very opportunistic behaviour at the very early phase of the process. This was supported by the findings of Stientra et al., that telcos followed 'herd-behaviour' in the early internationalisation phase, but later started to make more strategic choices.

However, these studies were focused largely on the internationalisation of the largest telcos and have not included some of the most recent developments in the industry. Thus, more information on the process is still needed. As mentioned earlier, due to the capital intensity of the industry, telcos from smaller countries faced even greater challenges in their internationalisation. Whereas the largest telcos from the largest countries started to compete in global markets and had sufficient resources to enter several of the largest markets with FDIs, SMOPEC telcos were forced to look for alternative strategies. These largest telcos

Figure 2. Conceptual Framework: Internationalisation strategies and factors influencing them



were companies such as AT&T, Britisth Telecom, Deutche Telekom, France Telecom/Orange and Japanese NTT, and new global telecommunications operators such as WorldCom, Vodafone and Hutchison (D'Aveni, 2002; Heng & Low, 1990; Li & Whalley, 2002; Whalley, 2004). Many studies reported that the telecommunications service sector started to globalise (Ramamurti, 2000) following the earlier development of the telecommunications manufacturing sector, and this resulted in rapid consolidations of its players (Sirel & Wavermann, 2000). For example, Whalley (2004) predicted that in the mobile communications industry two types of operators would emerge: the global flag-ship operators such as Vodafone and Hutchison with global brands, many centralised operations and economies of scale advantages; and smaller regional operators.

How have telcos from smaller countries been able to reposition themselves in this evolving industry value network and survived? Have they implemented niche strategies as suggested by some theories based on SMOPEC research (Benito et al. 2002; Dick and Merrett, 2007), or have they entered into strategic alliances as suggested by others (Merrett, 2002; Ramamurti & Sarathy, 1997); or have they been forced to withdraw from these types of very capital-intensive industries altogether (Morkel & Osegowitsch, 1999)? Another little researched area from the internationalisation process perspective is the development of joint-ventures (JVs) in the telecommunications industry. Studies have found that due to some remaining regulations in the industry, telcos in general have been forced to form JVs with local partners in many country markets (Granstrand, 1994). However, whether these types of entry modes have been utilized by SMOPEC telcos is still an under researched area. What impact these factors have had on the internationalisation strategies of SMOPEC telcos and their ability to identify/develop competitive advantages to overcome the smallness of their domestic markets, are also questions that this study is investigating. To illustrate process of internationalisation of telcos from SMOPECs, the findings from four case studies are discussed as follows.

CASE EXAMPLES ON THE INTERNATIONALISATION OF SMOPEC TELCOS

The case companies used here to illustrate the internationalisation processes of telcos from SMOPECs and on the overall development of the industry value network, originate from three different continents: Sonera Oyj (Sonera) from Finland, Telia AB (publ) (Telia) from Sweden, Singapore Telecommunications Limited (SingTel) from Singapore, and Telstra Corporation Limited (Telstra) from Australia[3].

All of the aforementioned telcos were established in the early years of the industry, more than a century ago. For many decades they enjoyed protection from competition serving their home markets as national monopolies in most or all of their business areas. First international activities included co-operation with other national telcos from other countries. However, this was limited mostly to bilateral and multilateral interconnection negotiations and other friendly co-operation activities.

The first phase of internationalisation with significant outward operations did not start until the 1970s and 1980s. However, when this occured their internationalisation patterns deviated from those predicted by traditional theories, especially with regards to market strategies. The first international operations were management contracts and consulting projects, but not in countries within a small psychic distance, but in developing countries, often even outside their home continents. For instance, Finnish Sonera established operations in Africa and South East Asia; Swedish Telia in Africa and Asia; Singaporean SingTel in the Middle East and some developing

African and Asian countries, and Australian Telstra in the Middle East and developing Asian countries. That is, mainly target markets with long cultural, economic and/or geographical distance from home markets[4]. Interestingly, the case companies adapted their product strategies. Instead of trying to enter international markets with their asset-based/location-bound services, as predicted by research findings on these types of service companies, they operated through consulting-projects and management contracts by 'exporting' their personnel. In this regard their product and operation strategies followed the incremental patterns that the traditional process theories for manufacturing companies suggested. This was against general predictions of an early entry mode in an asset-based-location-based service industry.

In the second phase, in the 1990s, the case study companies entered international markets with more committed operation modes, such as JVs and, in some rare cases, in full-ownership modes. It can be argued that during this phase the target market selections and product strategies were very opportunistic with no clear patterns. For example, Telstra acquired interests in an Indonesian fixed-line business, and also had some operations in India and Vietnam, but they also established a full-service subsidiary in New Zealand. Soon they also acquired full ownership of a mobile operator in Hong Kong. Telia invested in mobile and fixed-line operators in the Baltic countries (Estonia, Latvia, Lithuania), and in many Latin American and African countries, but also established/acquired subsidiaries in Denmark and Finland. Sonera invested in JVs in both mobile and fixed-line operators in the Baltic countries, and in datacom and mobile operations in Russia. In addition it invested in mobile operators in Hong Kong, Hungary, Lebanon and Turkey. SingTel acquired shares in a full-service operator in Philippines, mobile operators in Belgium and Norway, and in cable-TV operators in Sweden and the U.K. Soon they also entered into equity partnerships with mobile operators in Thailand and India.

With regard to their market strategies the case companies' internationalisation patterns varied significantly from those suggested by traditional internationalisation process theories. Most entries of the case companies were not to neighbouring developed countries with small psychic distance, but to culturally very different, economically less developed and, in many cases, also geographically distant countries.

It could be argued that in both phase one and two the market strategies were very opportunistic, based on the need to achieve first mover advantages in an industry with network externalities, oligopolistic structure and the overall competitive situation. What contributed to the deviations in the market strategies of the case companies was that they had a competitive advantage in target countries due to their technical and marketing knowledge, but this advantage applied mostly in developing markets, rather that in other developed markets. However, the high capital intensity and host government regulations also created some limitations. As discussed, some of these challenges may have been significant especially for telcos from SMOPECs. This was reported to be one reason why the case companies were willing to remain as minority shareholders in many of their international JVs, a position that many of the more dominant industry operators seemed to try to avoid and overcome.

It must also be noted, though, that at the end of phase two the case companies also had implemented some relatively aggressive and significant entries to some large and developed countries, such as SingTel's acquisition of Optus in Australia, Sonera's acquisitions of a minority share of a US-based mobile operator and an expensive 3G-license in Germany, and the significant investments by Telia, Telstra and SingTel in large regional data networks. This was partly caused by the 'follow the herd' phenomen in this oligopolistic industry, further intensified by domestic push forces and pressures to grow mainly, but not exclusively from the financial markets.

In addition, during this high growth phase in the industry some of the case companies established specialised subsidiaries with objectives to achieve a global market leadership in their own niche businesses, such as Sonera's Zed and Telia's Speedy Tomato in the mobile portal business, and Sonera's Smart Trust in the mobile payment sector. It can be further argued that at some level they tried to follow similar strategies to born global companies. However, these operations were not very successful and most of them have either been divested or terminated. While these events can be partly attributed to changes in market sentiment, difficulties in trying to combine very different organisation and market strategies within one group of companies was also an important factor. In other words, the parent companies followed multinational organisation strategies in which units in very different host countries were operating relatively independently, whereas the 'born global' companies tried to follow more global organisation structures. Further, it seemed that within an oligopolistic industry in which competition among telcos intensified after borders had opened, other telcos were reluctant to buy content from their potential competitors' subsidiaries. Clearly the search for new positons in the evolving value network was a great challenge for the case companies.

By the end of 2000 and early 2001, the general market sentiment deteriorated significantly. This resulted in changes in the internationalisation of the case companies. Thus, the third phase of their internationalisation was in clear contrast to the aggressive and opportunistic strategies in phases one and two. In many areas the processes were reversed, as the case companies started to focus on their core competences and on markets close to home, resulting in divestments of several of their earlier investments. Most new investments in distant markets were frozen. These de-internationalisation developments were different from the more deterministic and linear processes that most traditional process models suggested[5]. With regard to market strategies the focus was now on the home continent, or even the more narrowly defined home region, such as Northern Europe for Telia. The era of globalisation was over, at least for now, and the case companies became regional. Also, their product strategies became more focused; for example, on mobile communications in some new high growth markets, although the case companies still remained as integrated full-service operators in their home and close neighbouring markets, rather than becomng pure niche companies with a focus on just one business area.

In the last phase, starting in 2002/2003, the case companies started to slowly look for new growth opportunities, but this time growth was focused mostly on new investments in their own region, in neighbouring countries, and to increase their ownership shares in their existing investments. It can be argued that their organisation structures developed from multinational to more transnational ones. They started to look for and gain synergies across their regional country organisations. This became possible as differences between different country organisations declined and operational and managerial control became easier due to increased ownership shares. Moreover, the overall internationalisation patterns became more gradual, and closer to the patterns suggested by traditional process theories. Examples of this phase were Telia's and Sonera's merger to form a multi-country telco in Northern Europe, and SingTel's closer integration of many of its activities with its largest subsidiary, Optus in Australia, and also closer integration between its' other regional JVs. However, in spite of this closer integration, the individual country organisations of each case company maintained their domestic company names and brands; that is, in many areas multinational strategies prevailed, instead of them moving towards more integrated and global-type of strategies.

It is also notable that throughout the four phases of internationalisation discussed previously, in their business-to-business (B2B) operations the case companies followed internationalisation patterns comparable to the ones reported in earlier studies of B2B-services (Aharoni, 2000; Roberts, 1998). That is, they established offices in other developed countries, in neighbouring countries with both small cultural and geographical distance, but especially in the leading markets in their continent and/or around the world. For instance, they established offices in London and other large cities in Europe, SingTel in most major Asian business cities, and Telstra in many of them, and all of them also had offices in the U.S. In addition, they joined in alliances with other telcos to be able to provide even more comprehensive global services to their global MNC-customers. The process, with regard to operation strategies, was relatively rapid. This is common among business services with follow-the customer strategies, as uncertainties and risks to internationalisation are rather low. This is both because of the relatively small size of the offices and, as for the case companies, the first entries were made to serve their domestic industrialised customers who had internationalised earlier. On the other hand, with regard to market strategies, the B2B-activities followed much more traditional patterns, also supporting earlier theories on B2B-services. It must be noted, though, that for SMOPEC telcos there are relatively fewer MNCs to follow, as the head quarters of most of the largest MNCs in the world originate in the largest economies rather than in SMOPECs. Thus, it could be argued that the importance of international B2B customers has been smaller for SMOPEC telcos than for the most globalised large country-originated telcos. So far the share of the revenues for the case companies from their international B2B operations have been moderate when compared to their B2C and domestic B2B operations[6].

When analysing factors that have contributed to these developments, it can be seen that ***industry specific factors*** such as deregulation/regulation developments, the oligopolistic nature of the industry (limiting opportunities in each country market), economies of scale advantages[7], technological developments, the growth pace of the industry, and the nature of the products (i.e. asset-specific/capital-intensive services consisting of complex systems and physical networks), have caused their internationalisation strategies to be rapid and irregular. This supports the findings of Sarkar et al. (1999) discussed earlier. Rapid industry growth together with limited opportunities created high real and perceived 'first mover advantages'. Thus telcos sought country markets with limited barriers to entry almost independent of their geographical location. These results also support Ramamurti's (2000) findings of 'first mover advantage' in the industry, and Stientra et al.'s (2004) findings of 'herd-behaviour' mentioned earlier.

However, the high initial capital investments that were required in telecommuinications networks created additional challenges to the case study companies, as was predicted. With their limited resources and due to the relatively high risks involved, they were not able to enter large markets rapidly, at least not several large markets simultaneously. This varied from the strategies of the largest national telcos in the industry and of the more specialised companies such as Vodafone in mobile communications and Equant in data communications. For SMOPEC telcos entries to large developed markets seemed to be rare exceptions, and most of these few entries resulted in de-internationalisation and/or divestment decisions at a later phase of the process.

Operation strategies of the case companies were based mostly on committed operation modes, as expected in network industries, but there was also interesting evidence of adaptations in their product strategies to make it possible to use less committed operation modes in the early phases of the internationalisation process. This strategy followed more of the processes suggested by the traditional intenationalistaion theories on manufacturing companies or hard-services, a finding that could provide an example of an alternative early phase entry strategy for a telco from a smaller country, when compared to more aggressive international entries by larger service MNCs in these types of industries.

Also, the use of strategic alliances and JVs was common, following the expectations for a SMOPEC telco in a capital-intensive industry. However, it needs to be emphasized that the role of non-equity alliances diminished significantly over the process of internationalisation. It seems that even for telcos from SMOPECs the challenges of entering and maintaining an alliance with another telco in a situation in which competition is rapidly globalising, became greater than the possible benefits and synergies from the alliance. On the other hand, throughout the process JVs have been the most common international operation mode for the case companies. Unlike many large telcos, the case companies seemed to be relatively compliant as minority owners in many of these JVs. All of them emphasized the importance of owning a significant share of a JV and being the largest telco owner, but in many cases the ownership shares of 20 - 40 % were perceived to be sufficient. It could be argued that this willingness to work together with local partners, instead of aiming actively for a dominant role may have contributed to the attractiveness of these companies as JV partners. This difference in approaches may also relate to differences in their organisation strategies: for a global company the dominant position is important in order to gain control of the operations and to align them with a more centralised model. This was rarely a requirement from the SMOPEC telcos. In addition, as the telco sector is often perceived to be very strategic for a country and governments play an important role, political strategies are often a very important factor in the JV negotiations. In this type of environment SMOPEC telcos were often perceived to be less threatening and the power balance to be more even with local partners than would be the case with telcos from the largest economies in the world. It could be argued then, that this can be a competitive advantage for SMOPEC telcos when entering foreign markets, especially in many developing countries.

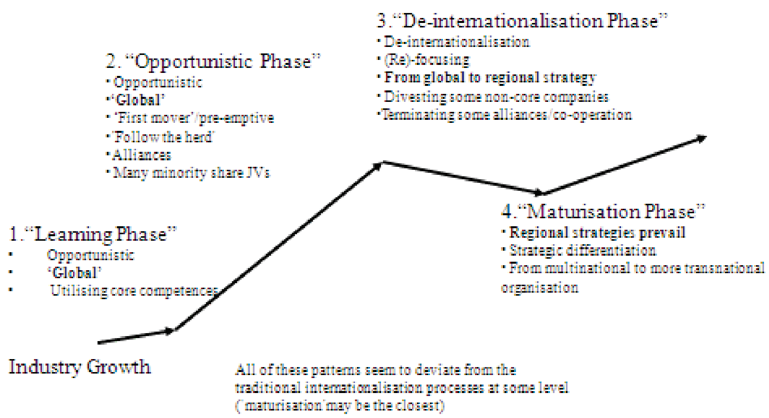
With regard to their product strategies, the case companies tried to implement global niche strategies, but unlike for many SMOPEC firms, these were not found to be a successful way to internationalise. It could be argued that these strategies did not fit well with the overall strategies of the case companies.

It can be argued that telcos have followed a different path in their internationalisation than, for example, large manufacturing companies or 'born globals'. This supports the earlier studies of Sarkar et al. (1999) and Stientstra et al. (2004). In addition, several features were emphasized by telcos from SMOPECs: adjustments in product strategies at the early phase of their internationalisation, emphasis on distant developing markets in their market strategies, emphasis on minority JVs, and all implemented multinational organisation strategies (e.g. no global brands) with some later transnational characteristics. A further interesting finding was a possible competitive advantage arising from the SMOPEC telcos being perceived as less threatening by the host governments and potential host country JV-partners than similar companies from large, developed countries. All this may have contributed to the belief that SMOPEC telcos have not implemented global organisation strategies. It could equally be argued that when moving onwards from multinational structures/strategies, a transnational organisation strategy is more feasible for SMOPEC telcos than a global strategy; given that, the role of the domestic market is perceived as less important and dominant for the whole company[8]. As such, this may create a situation in which it is relatively easier/more natural for a SMOPEC telco to allow its country organisations to participate in the decision making and development processes of the whole company. Some early findings of this type of behaviour were observed in our cases, although this issue requires furter investigation[9].

Furthermore, the findings reveal that there have been four separate phases in which the case companies have internationalised: *Learning Phase*, *Opportunistic Phase*, *De-internationalisation Phase*, and *Maturisation Phase* (Laanti, 2008). This is demonstrated in Figure 3. This identification of separate phases in the process is different from most traditional and deterministic theories, and also adds to the existing research on services. It could be argued that the growth pace of the industry combined with some pressures from the financial markets have been significant factors contributing to these phases. During the early phases 'herd-behaviour' was relatively strong, whereas during the later phases more strategic differentiation occurred, which supports the integration of strategic management concepts to the conceptual frameworks analysing internationalisation processes of the firm. This also supports the findings of Stienstra et al. (2004), discussed earlier, but overall this is a topic that requires further research.

Although the first two phases included some global aspirations and strategies, and attempts to implement global focused niche-based product strategies with some of their operations, the case companies have recently implemented much more regional strategies. They have remained mostly regional players with integrated product strategies benefiting from economies of scope advantages when providing fixed, mobile and data services, especially in their domestic markets, and they have also remained vertically integrated with regards to their network ownership and service operations (rather than becoming virtual network operators). It needs to be noted that some researchers have argued that even the most globalised telcos, such as Vodafone and Hutchison, still generate a large part of their revenues from one continent (Rugman, 2003), and also seem to follow multinational strategies in some of their operations, such as network purchases (Whalley, 2004). That is, although there are differences in this perspective between telcos from SMOPECs and the largest and most 'global' telcos in the industry, all telcos still fall behind in globalisation developments when compared to most other companies in the industry value network. These observations support the findings of Proff (2002) and Rugman (2003) that inspite of general development from domestic to more international markets, many industries are still regional rather than global. At one extreme are global manufacturers such as Nokia and Ericsson (Rugman, 2003; Steinbock, 2003), and born global companies such as many global softaware companies (Bell, 1995). At the other end are many retailing companies, which

Figure 3. Phases in telco internationalisation (adapted from Laanti et al (2006))



have started to internationalise, but are still more regional rather than global in their nature (Rugman & Girod, 2003). Pure B2B services seem to follow the traditional process in their internationalisation, but they often do this more rapidly and end up to be 'global' companies (Roberts, 1999).

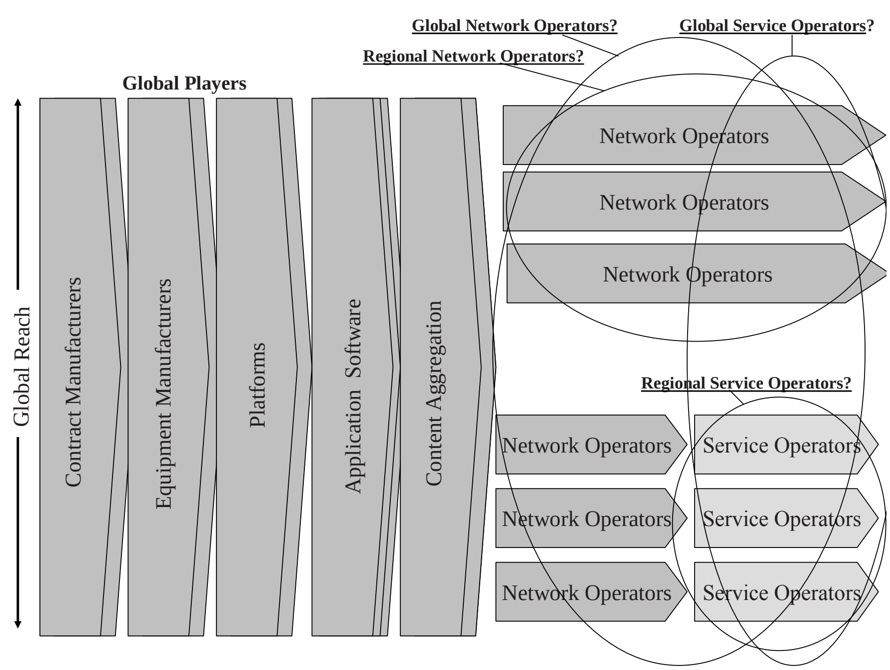
In Figure 4 (Global Industry Value Network) these different approaches to global and regional strategies between different companies within the telecommunications industry, and the vertical integration/disintegration within and across the value chains, are illustrated. Different strategic options for telcos are highlighted in the right side of the figure: global network operators, global service providers, regional network operators and regional service providers. As discussed, these developments seem to indicate that although some of the largest telcos have been implementing global strategies at some level, telcos from SMOPECs follow mostly regional network operator strategies. Some early tries to implement global service provider (niche) strategies or to diversify their operations to global application software or content aggregation strategies, have not been very successful. It remains to be seen how this situation will develop in the future.

CONCLUSION AND FUTURE RESEARCH OPPORTUNITIES

Telcos' share of the domestic value chain has decreased significantly as new players have entered their domestic markets and value chains have disintegrated into value systems. On the other hand, convergence has also opened new opportunities in new business sectors, both in domestic and internatonal markets. Together this more intensive competition domestically, and these new opportunities have forced telcos to enter international markets.

Based on the research results on value networks in the telecommunications industry, earlier research on telco internationalisation, and this study on the internationalisation of SMOPEC telcos, it can be argued that telcos have followed different

Figure 4. Global industry value network



internationalisation strategies to telecommunications equipment manufacturers or ‘born globals’ in the industry. The reasons for this can be found in their different business logic and in the characteristics of the sector, resulting in exposure to different types of challenges when compared to the manufacturing sector or ‘born globals’. For example, deregulation/regulation, the greater role of governments, network externalities, the different nature of economies of scale advantages, and the high capital intensity of the sector create unique challenges for telcos. All this has resulted in a unique competitive environment, with some oligopolistic structures. In addition, it was illustrated that telcos from SMOPECs face their own specific challenges, resulting in alternative patterns of internationalisation, but they have been able to develop some useful competitive advantages against their larger competitors.

It is evident that the development of value networks in the telecommunications industry or the ICT-industry is still ongoing. Interesting questions remain with regard to telcos’ internationalisation in the future: what will be the successful strategies, business models, and optimal positions within the industry value network for telcos more generally, and for telcos from SMOPECs particularly?

These questions will provide some future research avenues. First, more empirical and longitudinal studies are required to analyse market strategies of telcos, that is, horizontal (geographical) integration in the value network. These studies could analyse if regionalisation developments are just another phase in the process and if global strategies will prevail in the future. Second, this question leads to a related research topic. As mentioned, the different development levels between different country markets caused organisation strategies of most telcos (and all SMOPEC telcos) to be multinational rather than global. However, it would be interesting to investigate how the situation changes when the industry matures further and the development levels between country markets is closer than at present. That is, would this result in more consolidations of telcos across national borders and more integrated organisation strategies?

Third, the value chain and value network concepts have been useful to analyse operation strategies of telcos, especially their vertical integration vs. disintegration along the value chain. Now some network independent service providers have emerged in markets, but so far they have not been able to succeed against more integrated traditional telcos. More longitudinal studies on this issue are required.

Fourth, with regards to their product strategies most telcos still seem to follow mostly horizontally integrated product strategies, especially SMOPEC telcos in their domestic markets. By doing this they resemble many large retail companies in their business logic. That is, they sell and package services developed by their international suppliers, and utilise economies of scope advantages. Future studies could analyse if these are the winning strategies in the future or will more focussed product strategies become more successful, as they may allow more opportunities to specialise and internationalise, even for companies with limited resources. Naturally, one of the most relevant and important questions for SMOPEC telcos is whether they are able to survive as independent organisations or will the long predicted consolidation developments eventually result in a few large telcos dominating the sector.

These and other questions with regards to optimal positioning in the globalising ICT industry seem to remain relevant to researchers, managers of internationalising telcos, to other industry players, and to policy makers. In investigating these questions the concept of value network and the internationalisation process models can provide a useful analytical framework.

ACKNOWLEDGMENTS

The authors want to acknowledge the key interviewees of each case company for their valuable support for the project. The authors want also to thank the following organisations for providing funding and support to the project: The Foundation for Economic Education, The Finnish Cultural Foundation, HPY Research Foundation, and International Telecommunication Union (ITU).

The early version of this paper has been presented in ITU 2006 Telecom World Conference (4th- 8th December) in Hong Kong. The authors have a permission from International Telecommunications Union to publish the developed version also as a book chapter.

REFERENCES

Aharoni, Y. (2000). The role of reputation in global professional business services. In Y. Aharoni & L. Nachum (Eds.), *Globalization of Services: Some implications for theory and practice*. London: Routledge.

Bartlett, C. A., & Ghoshal, S. (1992). *Transnational Management: Text, Cases and Readings in Cross-Border Management* (3 ed.). Irwin: McGraw-Hill.

Bell, J. (1995). The internationalization of small computer software firms: a further challenge to "stage" theories. *European Journal of Marketing, 29*(8), 60-75.

Benito, G. R. G., Larimo, J., Narula, R., & Pedersen, T. (2002). Multinational Enterprises from Small Economies. *International Studies of Management & Organization, 32*(1), 57-78.

Benito, G. R. G., & Welch, L. S. (1994). Foreign Market Servicing: Beyond Choice of Entry Mode. *Journal of International Marketing, 2*(2), 7-27.

Bilkey, W., & Tesar, G. (1977). The Export Behavior of Smaller-sized Wisconsin Manufacturing Firms. *Journal of International Business Studies, 8*(1), 93-98.

Boddewyn, J. J., Halbrich, M. B., & Perry, A. C. (1986). Service Multinationals: Conceptualization, Measurement and Theory. *Journal of International Business Studies, 17*(3), 41-57.

Brouthers, K. D., & Brouthers, L. E. (2003). Why Service and Manufacturing Entry Mode Choices Differ: The Influence of Transaction Cost Factors, Risk and Trust. *Journal of Management Studies, 40*(5), 1179-1204.

Cave, M., & Waverman, L. (1999). The transformation of international telecommunications. *Business Strategy Review, 10*(1), 21-27.

Cavusgil, S. T. (1984). Differences Among Exporting Firms Based on Degree of Internationalization. *Journal of Business Research, 12*(2), 195-208.

Clark, T., & Knowles, L. L. (2003). Global myopia: globalization theory in International Business. *Journal of International Management, 9*(4), 361-372.

Clark, T., Rajaratnam, D., & Smith, T. (1996). Toward a Theory of International Services: Marketing Intangibles in a World of Nations. *Journal of International Marketing, 4*(2), 9-28.

Contractor, F. J., & Kundu, S. K. (2000). Globalization of hotel services: an examination of ownership and alliance patterns in a maturing service sector. In Y. Aharoni & L. Nachum (Eds.), *Globalization of Services: Some implications for theory and practice*. London: Routledge.

Contractor, F. J., & Lorange, P. (2002). The growth of alliances in the knowledge-based economy. *International Business Review, 11*(4), 485-502.

D'Aveni, R. A. (2002). Competitive Pressure Systems. *MIT Sloan Management Review, 44*(1), 39-49.

Dick, H., & Merrett, D. (2007). *The Internationalisation Strategies of Small Country Firms*. Cheltenham, UK: Edward Elgar.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization, 14*(6), 673-699.

Erramilli, M. K. (1990). Entry Mode Choice in Service Industries. *International Marketing Review, 7*(5/6), 50-62.

Erramilli, M. K., & Rao, C. P. (1993). Service firms' international entry-mode choice: A modified transaction-cost analysis approach. *Journal of Marketing, 57*(3), 19-38.

Fjeldstad, Ø., Becerra, M., & Narayanan, S. (2004). Strategic action in network industries: an empirical analysis of the European mobile phone industry. *Scandinavian Journal of Management, 20*(1/2), 173-196.

Fransman, M. (2002). Mapping the evolving telecoms industry: the uses and shortcomings of the layer model. *Telecommunications Policy, 26*(9/10).

Gabrielsson, P., & Gabrielsson, M. (2004). Globalizing internationals: business portfolio and marketing strategies in the ICT field. *International Business Review, 13*(6), 661-684.

Glachant, J.-M. (2002). Why Regulate Deregulated Network Industries? *Journal of Network Industries, 3*, 297-311.

Granstrand, O. (1994). Conclusions and Speculations. In E. Bohlin & O. Granstrand (Eds.), *The Race to European Eminence - Who are the coming Tele-Service Mulinationals?* (pp. 523-542). Amsterdam: Elsevier Science.

Häikiö, M. (2001). *The History of Nokia Oyj, Part III: Globalisation 1992–2000 (in Finnish).* Helsinki: Edita Oyj.

Heng, T. M., & Low, L. (1990). Towards greater competition in Singapore's telecommunications. *Telecommunications Policy, 14*(4), 303-314.

Javalgi, R. R. G., Griffith, D. A., & White, D. S. (2003). An empirical examination of factors influencing the internationalization of service firms. *Journal of Services Marketing, 17*(2), 185-201.

Johanson, J., & Vahlne, J.-E. (1977). The Internationalization Process of the Firm - a Model of Knowledge Development and Increasing Foreign Market Commitments. *Journal of International Business Studies*, *8*(1), 23-32.

Johanson, J., & Wiedersheim-Paul, F. (1975). The Internationalization of the Firm: Four Swedish Cases. *Journal of Management Studies, 12*(3), 305-322.

Kirpalani, M., & Luostarinen, R. (1999, Dec 14). *Dynamics of Success of SMOPEC Firms in Global Markets.* Paper presented at the EIBA Annual Conference, Manchester.

Knight, G. A. (1999). International services marketing: review of research 1980-1998. *Journal of Services Marketing, 13*(4/5), 347-360.

Knight, G. A., & Cavusgil, S. T. (1996). The Born Global firm: a challenge to traditional internationalization theory. *Advances in International Marketing, 8*, 11-26.

Laanti, R. (2008). *Internationalisation Strategies of Companies in Service Industries: A Study of National Telecommunication Operators from Small and Open Economies (SMOPECs).* Unpublished PhD, The University of Adelaide, Adelaide.

Laanti, R., Gabrielsson, M., & Gabrielsson, P. (2007). The globalization strategies of business-to-business born global firms in the wireless technology industry. *Industrial Marketing Management, 36*(8), 1104-1117.

Laanti, R., McDougall, F., & Baume, G. (2006). *Stages Or Leaps? - Market Strategies Of Telecommunication Operators From Small And Medium Sized Countries.* Paper presented at the 48th Annual Meeting of the Academy of International Business, Beijing.

Levitt, T. (1983). The Globalization of Markets. *Harvard Business Review, 61*(3), 92-101.

Li, F., & Whalley, J. (2002). Deconstruction of the Telecommunication Industry: From Value Chains to Value Networks. *Telecommunications Policy, 26*(9/10), 451-472.

Liebowitz, S. J. (2002). *Re-thinking the network economy: the true forces that drive the digital marketplace*. New York: AMACOM.

Liesch, P. W., Welch, L. S., Welch, D., McGaughey, S. L., Petersen, B., & Lamb, P. (2002). Evolving Strands of Research on Firm Internationalization: An Australian-Nordic Perspective. *International Studies of Management & Organization, 32*(1), 16-35.

Lovelock, C. H., & Yip, G. S. (1996). Developing global strategies for service businesses. *California Management Review, 38*(2), 64-86.

Luostarinen, R. (1979). *Internationalization of the Firm*. Helsinki: Helsinki School of Economics.

Luostarinen, R. (1994). *Internationalization of Finnish Firms and their Response to Global Challenges - Research for Action*. Forssa: WIDER Report. The United Nations University/WIDER.

Maitland, E., & Nicholas, S. (2002). Preface: Modeling Multinationals from Small, Open Economies. *International Studies of Management & Organization, 32*(1), 3-15.

Majkgård, A., & Sharma, D. D. (1998). Client-Following and Market-Seeking Strategies in the Internationalization of Service Firms. *Journal of Business-to-Business Marketing, 4*(3), 1-41.

McDougall, P., Shane, S., & Oviatt, B. (1994). Explaining the Formation of International New Ventures: The Limits of Theories from International Business Research. *Journal of Business Venturing, 9*, 469-487.

Merrett, D. T. (2002). Australia's Emergent Multinationals. *International Studies of Management & Organization, 32*(1), 109-135.

Morkel, A., & Osegowitsch, T. (1999). The international competitiveness of Australian business. In G. Lewis, A. Morkel, G. Hubbard, S. Davenport & G. Stockport (Eds.), *Australian and New Zealand Strategic Management: Concepts, Context and Cases*. Sydney: Prentice Hall Australia Pty Ltd.

OECD. (2007). *Glossary of Statistical Terms*. Retrieved October 13, 2007, from http://stats.oecd.org/glossary/search.asp

Ohmae, K. (1989). Managing in a Borderless World. *Harvard Business Review, 67*(3), 152-161.

Peppard, J., & Rylander, A. (2006). From Value Chain to Value Netork: Insight for Mobile Operators. *European Management Journal, 24*(2-3), 128-141.

Porter, M. E. (1985). *Competitive advantage: creating and sustaining superior performance*. New York: Free Press.

Proff, H. (2002). Business unit strategies between regionalisation and globalisation. *International Business Review, 11*(2), 231-250.

Ramamurti, R. (2000). Risks and Rewards in the Globalization of Telecommunications in Emerging Economies. *Journal Of World Business, 35*(2), 149 - 170.

Ramamurti, R., & Sarathy, R. (1997). Deregulation and globalization of airlines. *International Trade Journal, 11*(3), 389-432.

Rennie, M. (1993). Global Competitiveness: Born Global. *The McKinsey Quarterly, 4*, 45-52.

Roberts, J. (1998). *Multinational Business Service Firms: The Development of Multinational Organisational Structures in the UK Business Services Sector*. Aldershot, England: Ashgate Publishing.

Roberts, J. (1999). The Internationalisation of Business Service Firms: A Stages Approach. *Service Industries Journal, 19*(4), 68-88.

Rugman, A. M. (2003). Regional strategy and the demise of globalization. *Journal of International Managment, 9*(4), 409-417.

Rugman, A. M., & Girod, S. (2003). Retail Multinationals and Globalization: The Evidence is Regional. *European Managment Journal, 21*(1), 24-37.

Sabat, H. K. (2002). The evolving mobile wireless value chain and market structure. *Telecommunications Policy, 26,* 505-535.

Sarkar, M. B., Cavusgil, S. T., & Aulakh, P. S. (1999). International Expansion of Telecommunication Carriers: The Influence of Market Structure, Network Characteristics, and Entry Imperfections. *Journal of International Business Studies, 30*(2), 361-382.

Shapiro, C., & Varian, H. R. (1999). *Information rules: a strategic guide to the network economy.* Boston, Mass.: Harvard Business School Press.

Shy, O. (2002). *The economics of network industries.* Cambridge: Cambridge University Press.

Sirel, E., & Wavermann, L. (2000). 3G Mobile: The Push to Wireless Data across Europe. *Business Strategy Review, 11*(2), 67-71.

Steinbock, D. (2003). Globalization of wireless value system: from geographic to strategic advantages. *Telecommunications Policy, 27*(3/4), 207-235.

Stienstra, M., Baaij, M., Van den Bosch, F., & Volberda, H. (2004). Strategic Renewal of Europe's Largest Telecom Operators (1992-2001): From Herd Behaviour Towards Strategic Choice? *European Managment Journal, 22*(3), 273-280.

Welch, L., & Benito, G. R. G. (1996). De-internationalization. In *Working papers: International Business Enterprises.* Kingswood, N.S.W.: Department of Marketing, University of Western Sydney.

Whalley, J. (2004). Flagship firms, consolidation and changing market structures within the mobile communications market. *Telecommunications Policy, 28*(2), 161-175.

Yip, G. S. (1989). Global Strategy…In a World of Nations? *Sloan Management Review, 31*(1), 29-41.

Yip, G. S. (2000). Global Strategy in the Internet Era. *Business Strategy Review, 11*(4), 1-14.

KEY TERMS*

Globalisation: The OECD (2007) defined globalisation from an economic perspective: "*The term globalisation is generally used to describe an increasing internationalisation of markets for goods and services, the means of production, financial systems, competition, corporations, technology and industries*". Whereas Clark and Knowles (2003) included a broader view in their definition: "*The process by which economic, political, cultural, social, and other relevant systems of nations are integrating into World Systems is called globalisation.*" It seems that when analysing the internationalisation of many service companies, especially in network industries such as the telecommunications industry, it is necessary to include this broader perspective of globalisation into the conceptual framework, as often political and cultural systems play a significant role in the internationalisation of services.

ICT Industry: The convergence of the telecommunications industry with the computing and broadcasting industries resulted in a broader definition of Information and Communications Technologies (ICT). As defined by OECD: "*…ICT sector refers to equipment and services related to broadcasting, computing and telecommunications, all of which capture and display information electronically.*" (in UN Social Economic Council's Report of the International Telecommunication

Union on information and communication technologies statistics, 2004).

Internationalisation Process: describes a dynamic process of a firm's entry to international markets. In most cases the longitudinal studies of international processes include analyses of firms' operation modes and market strategies. Most of the traditional internationalisation process models have emphasized the gradual and incremental nature of the firm's internationalisation (see e.g. Bilkey & Tesar, 1977; Cavusgil, 1984; Johanson & Vahlne, 1977; Johanson & Wiedersheim-Paul, 1975; Luostarinen, 1979, 1994).

Small and Open Economies (SMOPECs): Small and open economies include countries such as Austria, Denmark, Finland, Ireland, New Zealand, Portugal, Norway, Sweden, and Switzerland, who have integrated themselves with the world economy by lowering or eliminating their trade barriers (Benito et al., 2002; Kirpalani & Luostarinen, 1999; Maitland & Nicholas, 2002; Merrett, 2002). The broader definition includes also medium-size countries such as Australia and newly industrialised countries such as Hong Kong (Maitland & Nicholas, 2002). Although Australia is already a medium-sized country with regards to its population, its companies face similar challenges in their internationalisation than companies from other SMOPECs (Dick & Merrett, 2007; Liesch et al., 2002).

Telcos: Telco is the term for a telecommunication operator; that is, a company that provides telecommunications services such as fixed-line, mobile and data services for end-customers. Most of the traditional telcos have been government owned telecommunication companies and usually also national monopolies, or at least duopolies.

Telecommunications Industry: Consisted traditionally of the manufacturers and network operators in fixed, mobile and data communications businesses. More recently many other companies, such as contract manufacturers, service providers, and application and content providers have emerged to be important players in the industry (see also the definition of the ICT industry).

Value Chain: The concept of *value chain* has been widely covered in the strategic management literature, most notably by Porter (1985). A value chain arises from a company's activities and internal business processes to create value for its customers (Porter, 1985). This value chain of a company then belongs to an industry value chain or system (Porter, 1985).

Value Networks: The difference in the definition of value networks, when compared to the traditional definition of value chains, is that in a value network there are several entry and exit points, and that activities take place simultaneously instead of successively (Li & Whalley, 2002).

ENDNOTES

[*] Most of these definitions are used also in the wider study on the internationalisation of telcos that this paper is based on (Laanti, 2008).

[1] It must be noted, though, that in some countries the telecom equipment manufacturers have been independent actors also historically, although often very closely linked with national telcos.

[2] Telecommunications equipment manufacturers were actually the first companies in the industry who started to internationalise rapidly.

[3] These case data were collected as a part of a broader cross-border multi-case study of the internationalisation of SMOPEC telcos that was mentioned earlier (Laanti, 2008). Australia and small and newly industrialised Asian countries were included as SMOPECs, based on suggestions by Maitland and

Nicholas (2002). While it is acknowledged here that Australia is a medium-sized country with regards to its population size, its companies nevertheless face similar challenges in their internationalisation as companies from other SMOPECs (Dick & Merrett, 2007; Liesch et al., 2002).

4 It must be noted that many of these projects were based on governments' foreign aid to developing countries. Thus, the risks and uncertainties for the telcos were relatively limited. This issue seem to further emphasize the important role of governments in the internationalisation processes of telcos.

5 Note that some studies on internationalisation processes have also reported remarkable de-internationalisation activities (e.g. Welch & Benito, 1996).

6 Due the position of Singapore as a regional hub for many global MNCs SingTel seems to have had a better position compared to other SMOPEC telcos to target international B2B customers and it seems that it has been relatively successful in this perspective, at least at the regional level.

7 Instead of global scale of advantages, in network industries the most scale advantages occur at local or regional level.

8 For instance, SingTel's domestic operations are much smaller than their Australian operations.

9 Some evidence of successful transnational organisation strategies were observed in some of the case companies, such as the internationally diversified compositions of the board of directors and management teams; and the active search and import of ideas between different country organisations as well as from other country organisations back to the home country - a key characteristic of a transnational organisation.

Chapter XII
The Effect of Non-Market Strategies in the Mobile Industry

Zulima Fernández
Universidad Carlos III de Madrid, Spain

Belén Usero Sánchez
Universidad Carlos III de Madrid, Spain

ABSTRACT

This chapter analyzes the role of non-market actions (NMAs) in the competitive position of firms in the European mobile telecommunications industry. First, the bases and uses of NMAs are reviewed and classified. Second, the use of legal actions as NMAs to erode the competitive position of first-mover firms is examined. Third, the findings of an empirical study confirming the effectiveness of legal actions as competition tools in the European mobile telecommunications industry are analyzed. Lastly, some conclusions, future lines of research and practical recommendations are presented

INTRODUCTION

Over the last decade the European mobile telecommunications industry has experienced unprecedented growth. The industry has expanded from a mere 35 million users with an average penetration rate of 12 percent in 1996 to more than 450 million users with an average penetration rate of over 100 percent at the end of 2006. In fact, 17 European Union (EU) member states have now exceeded 100 percent penetration (European Union, 2007).

The spread of digital technologies and market liberalization is responsible for this rapid expansion. Not all operators, however, have benefited equally. The technology in the mobile telecommunications industry is available to all firms, and identifying and copying innovations introduced by competitors is easy. Although this situation should cause pioneer operators to lose their first-mover advantage rapidly, pioneers have been shown to maintain their advantage (Fernández and Usero, 2007). An analysis of other—presumably more effective—ways of competing for follower firms, then, is needed.

Firms can defend themselves against rivals with a variety of competitive actions such as product, pricing or advertising activities (Smith et al., 1992). These are all market actions, but non-market actions (NMAs) are available to firms too. NMAs target politicians and regulators, the media, citizens and courts. Lobbying government, managing public or stakeholder opinion, and taking legal actions against competitors or the government itself are all examples.

The use of NMAs–particularly legal actions–to obtain competitive advantages has become "a necessity, not an option, in business" (Shell, 2004: 19). NMAs serve to complement–even replace–more conventional actions that are designed to implement a firm's strategy (Yoffie and Bergenstein, 1985). Their use is especially effective when business models and technologies can be easily replicated or when many of the business opportunities are controlled by government (Baron, 1995), as is the case in the mobile telecommunications industry.

Evaluating the effectiveness of NMAs to erode pioneer advantage in this industry, then, is a topic that merits discussion. Legal actions such as filing lawsuits and making complaints to regulatory bodies are among the most typically used NMAs. It is, for example, common for operators without significant market power to complain to regulatory agencies about high interconnection charges or difficulties to install antennas, while dominant operators complain about the help they are required to provide other firms. Despite this, however, few empirical works rigorously study the impact of this type of NMA. Consequently, this chapter has two objectives:

1. To discuss the bases and uses of NMAs and highlight the main types employed in the mobile telecommunications industry.
2. To analyze the use of legal actions to erode the competitive position of first-mover firms in the mobile telecommunications industry.

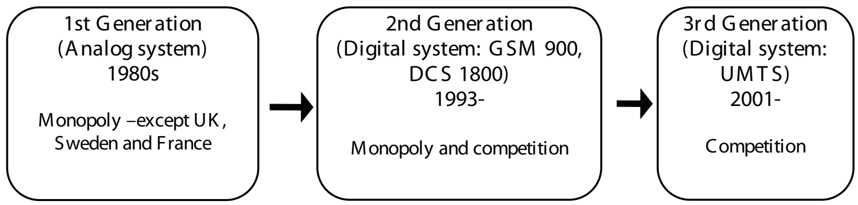
The chapter is organized as follows. First, the European mobile telecommunications industry is briefly described. Second, the effectiveness of NMAs as competition tools and their relevance to the mobile telecommunications industry are discussed. Third, the different types of advantages that legal actions bring are analyzed, along with the findings of an empirical study confirming the effectiveness of legal actions at eroding the pioneer advantage of European operators. The chapter concludes with some final thoughts and outlines some possible future lines of research.

THE EUROPEAN MOBILE TELECOMMUNICATIONS INDUSTRY

One of the reasons the mobile telecommunications industry has grown rapidly in Europe is the strong support given to it by the European Union (EU). The EU has made great efforts to establish a harmonized framework for the regulation of electronic communications services and networks, and associated facilities and services. The process began with the Green Paper on Mobile and Personal Communications (COM(94)145 final) and finished with Directive 2002/21/EC and another four specific Directives on a common regulatory framework for electronic communications networks and services.

This growth of the European market can be clearly traced through three stages that correspond to the technology used (Figure 1). In the 1980s almost all monopolies operating in the fixed telecommunications industry obtained the sole license to operate with analog technology. The United Kingdom and Sweden had a duopoly from

Figure 1. The evolution of mobile telecommunications systems



the beginning, with France reaching a similar state four years later. In the 1990s the introduction of digital technology (GSM-900, DCS-1800) made it possible to establish a European standardization system based on GSM technology. The resulting trans-European network overcame the problems of incompatibility that had existed with the analog systems. In addition, all countries opened their domestic markets to competition. The process, though, was not similar in all cases. In Sweden, Germany, Denmark and Portugal two or three licenses were granted from the beginning, eliminating the monopoly that had existed with analog technology. In Switzerland and Luxembourg, however, it took more than four years for the regulatory authorities to grant a second license.

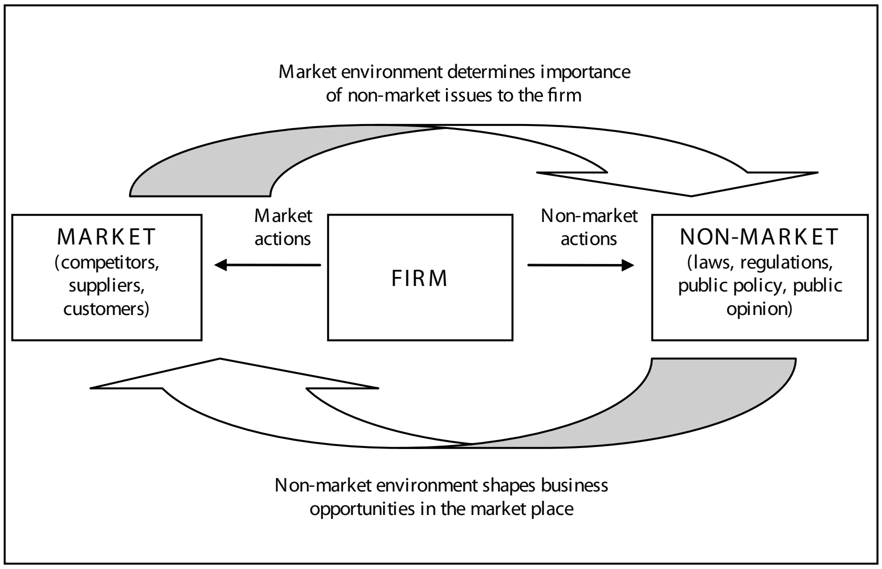
The licensing of third generation (UMTS) systems began in Europe in 2000. In countries like Spain and Sweden, the competitive tendering model was employed to encourage rapid deployment. In the UK, however, the government opted for a public tendering model. The huge revenue the UK government raised in this way encouraged many other European countries to adopt the same system, with the result that many firms have had their level of competitiveness damaged by the inflated prices they paid for UMTS licenses. Currently, the degree of UMTS implementation is still at a very embryonic phase. At present, there are 68 network operators in the EU, Norway and Switzerland offering digital mobile telephony services (GSM and, in most cases, UMTS) (see Appendix 1).

NON-MARKET ACTIONS AS COMPETITION TOOLS

Firms operate in both market and non-market environments. In market environments they relate to their competitors, suppliers and customers by undertaking actions such as launching new products, increasing capacity, cutting prices or running promotions. The non-market environment includes the interactions between the firm and other parties such as the public, the media and public institutions (Baron, 1995). The activity of the firm is limited by government regulation, while it is also often influenced by the media or the opinions and desires of stakeholders (e.g., employees or customers, among others). This is why firms attempt to influence norms, regulations and public policies in ways that favor their interests and–if possible–harm those of their competitors (Capron and Chatain, 2008). Similarly, firms take actions to shape public or stakeholder opinion. In short, non-market actions (NMAs) such as litigation, testifying at government hearings, lobbying, advocacy advertising, or constituency building are tools that firms can use to support their strategies.

NMAs complement–sometimes even replace–market actions taken to build competitive advantages; firms use them to attack or protect themselves from competitors. As with market actions, firms seek to shape or restructure the environment so that it is more favorable for their interests. NMAs may bring individual advantages for the undertaking firm (e.g., information not

Figure 2. Relationship between market and non-market environments



available to rivals) or advantages for the industry as a whole (e.g., adding trade barriers).

Non-market environments, then, shape the competition in the market, while firms' market actions modify non-market issues. In other words, both environments are inter-related (see Figure 2).

Firms may employ many types of NMA, each one with a specific objective and targeting a specific agent. Table 1 classifies the most important and common NMAs in the literature. Various types of actions are identified: information, financial contributions, constituency building and legal.

Table 1. Types of non-market actions (source: Adapted from Baron (1993) and Hillman and Hitt (1999))

TYPES OF NON-MARKET ACTION	DEFINITION
Information	Provide information to public decision makers, including measures such as: 1) Lobbying regulators, civil servants and politicians to influence their decisions. 2) Reporting research and survey results. 3) Testifying as experts in hearings or before other government bodies. 4) Supplying public decision makers with position papers or technical reports.
Financial contributions	Provide financial support, including measures such as: 1) Making financial contributions to politicians or to political parties. 2) Hiring personnel with political experience.
Constituency building	Seeks to gain voter, citizen and activist support or interest, including measures such as: 1) Grassroots mobilization of employees, suppliers or others linked to the firm. 2) Advocacy advertising, when a particular position is advertised to the public.
Legal	Formal complaints concerning allegedly infringed rights or dissatisfaction with the existing legislation, including measures such as: 1) Filing lawsuits against rivals or regulatory authorities. 2) Regulatory petitions.

Information actions and financial contributions are two types of NMA that give access to politicians and regulators (Schuler et al., 2002). Information actions seek to influence politicians and regulators by providing them with specific information about policy preferences or the consequences of decisions. Lobbying is the best known example. Firms also work with regulators by producing technical reports to advise the relevant authorities and by sending managers and technical experts to declare before them. In addition to this, firms try to access the policy-making process or exert influence over regulatory bodies by offering financial incentives to decision makers, or to the parties or institutions they belong to (Hillman and Hitt, 1999).

Likewise, constituency building encourages groups with similar interests to those of the firm to actively defend the firm's interests. Firms address different constituencies in order to encourage a particular behavior or position. Firms using NMAs have focused on public institutions such as legislators and regulatory agencies. Nevertheless, private interest groups (e.g., non-governmental organizations (NGOs) and activists) that attempt to influence the industry via the media and community mobilizations have now become the prime focus of these efforts. As a result, constituency building has become one of the most important NMAs used by firms.

Lastly, firms file lawsuits against their competitors or the public administration. They take this action to protect their interests and rights; these legal actions seek to handicap competitors or gain direct advantages (Baron, 1995). Figure 3 shows the main relationships among the agents in non-market environments.

NMAs are essential when opportunities are controlled by government (Baron, 1995), as well as when business models and technologies are easy to imitate. Because the mobile telecommunications industry has both characteristics (De Figueirido and Edwards, 2007), the use of various types of NMA is common. Table 2 details some examples of NMAs in this industry.

Although the effects of these NMAs have not been examined in great detail, some studies have begun to measure their impact. Bonardi (1999) examines how in the early 1980s British Telecom used its close relationship with the UK government to hinder the deregulation process in the telecommunications industry. And De Figueirido

Figure 3. Non-market actions and non-market agents

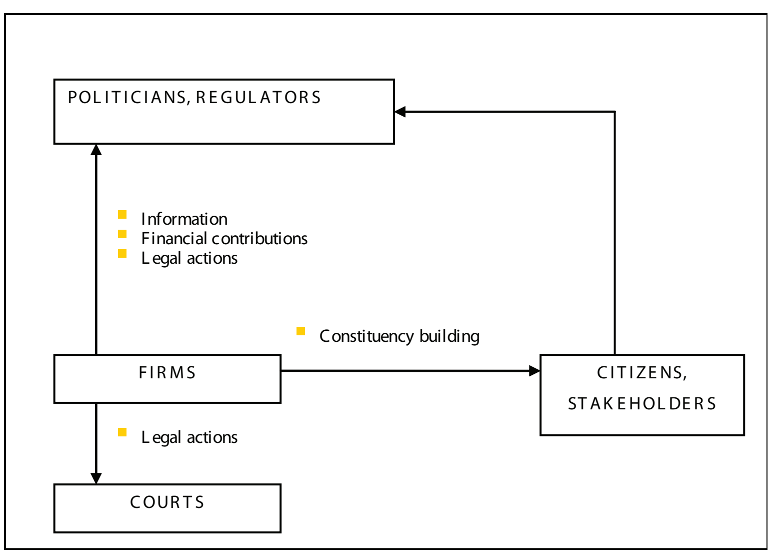


Table 3. First-mover advantages in the mobile telecommunications industry

NON-MARKET ACTION	EXAMPLE
Information	British Telecom tried to hinder deregulation of the British telecommunications industry (Bonardi, 1999)
Financial contributions	Private money in campaign contributions influences public policy outcomes (De Figueirido and Edwards, 2007)
Constituency building	Creation of a public opinion campaign designed to show that antennas posed no dangers to health
Legal	Telefónica Móviles España, S.A. embarked on legal actions against the Spanish regulatory authority (El País, 2006)

and Edwards (2007) look at how campaign contributions influence regulatory outcomes in the telecommunications industry. Specifically, they show how campaign contributions by incumbent firms and new entrants influence access prices for local loops of incumbents' telephone networks. This analysis also suggests that the effects of campaign finance on regulatory outcomes can be rapid.

Nonetheless, many other types of NMA exist that have not yet been studied. The next section focuses on one example–legal actions–that is particularly common in the mobile telecommunications industry.

THE EFFECT OF LITIGATION ON FIRST-MOVER ADVANTAGE IN THE MOBILE TELECOMMUNICATIONS INDUSTRY

Three types of first-mover advantage exist (Lieberman and Montgomery, 1988): technological leadership, pre-emption of scarce resources, and the creation of buyer-switching costs. First, technological leadership is gained by experience economies and R&D superiority. Second, pioneers can pre-empt scarce resources–such as favorable physical locations, inputs, plants and equipment or distribution and service systems–needed to develop the business. And lastly, switching costs serve to lock in customers. All these advantages make it more difficult for followers to compete with pioneers on a level playing field.

As mobile telecommunications is an equipment-based service, the advantages for pre-emption are significant (Bharadwaj et al., 1993). For example, the location of base stations poses problems for firms that enter the market later as fewer sites are available and local residents usually oppose placing them in towns and cities. Pioneers can also take better advantage of experience economies because they have more time to accumulate production and lower unit costs, even though all operators–regardless of when they enter the market–have the same technology. Non-portability of numbers, network economies and subsidized phones are the most important switching cost in the mobile telecommunications industry. Non-portability keeps the customer captive due to the inconvenience of having to change telephone numbers when switching to another operator. A similar effect occurs with the high price operators charge clients who wish to terminate their contract but maintain their telephone number. As mobile networks are highly compatible, network economies are likely to result from price discrimination inside and outside the network. Lastly, operators offer customers discounts on new phones at the end of their contracts. This offer may be the most important captive reason now that the problem of portability has been resolved (see table 3).

Some of these advantages are dependent on government and regulatory agencies. Time in monopoly has been shown to strengthen first-mover advantages (Fernández and Usero, 2007). It leads to an increase in customers, which favors the creation of network and scale economies and experience. Therefore, governments can reduce

Table 3. First-mover advantages in the mobile telecommunications industry (source: Adapted from Fernández and Usero (2007))

PRIMARY SOURCES	FIRST-MOVER ADVANTAGES IN THE MOBILE INDUSTRY
Pre-emption	Location of base stations Scale economies
Technological leadership	Experience economies
Switching costs	Number portability Network economies Discounts on the purchase of phones

first-mover advantages by allowing new operators to enter the market as soon as possible. Regulatory agencies can also ease or tighten restrictions on the establishment of base stations, depending on the degree of resistance from the public. They also have the power to force pioneers to allow competitors to use their base stations, and to impose portability.

Non-market actions (NMAs), then, are highly useful for pioneers as they can have an impact on government and regulators, and thus reinforce the first-mover advantage (Frynas et al., 2006). Followers use NMAs for similar reasons: to prevent pioneers from consolidating and improving their position. The main NMAs that followers have at their disposal are: providing public decision makers with information designed to encourage them to reconsider regulation, along with highlighting market power abuses by pioneers. These NMAs are undertaken by complaining to regulatory agencies, or as a last resort by taking the matter to court.

Pioneers and followers in the mobile telecommunications industry, then, use legal actions as NMAs for a number of reasons. These include the detection of allegedly illegal behavior by competitors, the need to counteract the effects of regulation, keenness to hinder the entry of new competitors, and the intention to harm rivals.

1. **Detecting allegedly illegal behavior by a competitor:** Although government agencies are normally responsible for supervising competitive restrictive conducts, private parties–either customers or competitors–can detect this kind of behavior more easily and may bring lawsuits against other firms for monopolistic behavior. An example is provided by the obligation of pioneers to offer followers the use of their base stations for a limited period of time (Laffont et al., 1997). Incumbents, though, commonly restrain followers from using their stations, thus forcing followers to complain to the regulatory agencies and courts. Pioneers themselves, however, also complain about possible abuses by followers when using shared resources.
2. **Influencing regulation:** Some operators in the mobile telecommunications industry have—significant market power—and work under more restrictive regulations than their competitors. These dominant operators, therefore, complain to the regulatory agencies and take legal actions to reduce the weight of regulation controlling interconnection prices and the rates competitors pay for the use of their installations or networks. In Spain, for example, Telefónica has filed more than 200 lawsuits against mobile interconnection charges, local loop unbundling, and universal service funding (El País, 2006).
3. **Hindering the entry or consolidation of new competitors:** In order to exploit the first-mover advantage for a longer period of time, pioneers may file lawsuits against new entrants. Their intention is to make the case last as long as possible, thus discouraging the entry of further competitors, or at least preventing them from competing on equal terms (Shell and Yao, 2000).
4. **Harming rivals:** The threat of legal actions distracts, delays, and imposes additional costs on rivals as the accused firm is forced

to divert time and resources from more productive activities to defend itself from the lawsuit. Empirical research has found that markets react negatively to the news of legal action and punish the defendant, but not the plaintiff (Koku et al., 2001). Litigation between firms causes the defendant significant losses of wealth (Bhagat et al., 1998). Most of these losses can be ascribed to the possibility of restrictions being placed on competitive behavior rather than the threat of fines (Bizjak and Coles, 1995).

Courts can be slow in reaching verdicts, so slow that decisions may even be handed down too late to help the plaintiff whose interests were abused by more powerful competitors. Litigation, though, can damage the defendant's reputation (Field et al., 2005). A claim of abuse of market power may adversely affect the public image of a firm and hence its sales. Firms that violate regulations suffer reputation loss. Indeed, nearly all the defendant's losses of wealth are caused by reputation loss rather than by penalties and litigation costs (Karpoff and Lott, 1999).

5. **Being criticized publicly by a competitor:** Firms sometimes make public–and baseless–attacks on their rivals, causing the accused firm to sue its competitors for damages.

In short, the evidence seems to point to pioneers and followers using legal actions to improve their competitive position. To test this hypothesis the authors studied the European mobile telecommunications industry with a database containing information on legal actions undertaken by pioneer and follower mobile phone firms from 1997 to 2000. The sample begins in 1997 because at that time there were at least two firms competing in the digital technology market in all of the countries. This makes it possible to study a group of pioneer firms as well as followers. The cut-off point was 2000 because in 2001 third generation (UMTS) licenses began to be granted.

The data on operators' legal actions were obtained by electronically searching the main general and business newspapers in the countries where they were working, along with several European trade journals from the telecommunications industry–all included in the Reuters database. Structured content analysis (Jauch *et al.*, 1980) was used to identify the legal actions of all the firms under study. The steps followed in the collection and sorting of news items are described in Appendix B (Usero and Fernández, 2005). The result of this meticulous analysis of news items was the identification of 109 legal actions performed by firms in the sample period (1997-2000).

A preliminary descriptive analysis of the data reveals that pioneers and followers in the EU have similar litigious conduct. As regulation is often the source of first-mover advantages in this industry, it would seem logical for followers to use legal actions to counteract them. Nevertheless, followers do not litigate more than pioneers do. Even though the data in table 4 might seem to indicate that follower firms are more likely to undertake legal actions than pioneers, these differences–with the exception of the data for 1999–are not statistically significant. Both types of firms take a similar number of legal actions. Followers seek to diminish pioneers' significant market power, and therefore constantly make demands to gain greater access to their rivals' resources (base stations, customers, etc.), to reduce interconnection charges, and to allow mobile number portability. Meanwhile, pioneers bring lawsuits against both competitors and regulatory authorities to defend their advantages.

A multiple regression analysis model was used to test whether the use of legal actions improves the competitive positions of pioneers and followers. The model includes two independent variables: pioneer legal actions in a particular country and year, and follower legal actions in a

Table 4. Legal actions mean test

	Mean	$Mean_{1997}$	$Mean_{1998}$	$Mean_{1999}$	$Mean_{2000}$
Mean (Legal $actions_{Followers}$) (*Standard Deviation*)	0.308 (0.568)	0.276 (0.527)	0.410 (0.594)	0.323 (0.584)	0.250 (0.556)
Mean (Legal $actions_{Pioneers)}$ (*Standard Deviation*)	0.205 (0.502)	0.206 (0.491)	0.256 (0.548)	0.102 (0.306)	0.279 (0.619)
p-value*	0.018	0.286	0.112	0.005	0.641

** P-values correspond to these mean tests: Mean (Legal $actions_{Followers}$) - Mean (Legal $actions_{Pioneers)}$) > 0*

particular country and year. It also incorporates various control variables that may affect the erosion of pioneer advantage: months in monopoly, order of entry, market penetration and industry growth. The measurement of variables is included in Appendix C.

The estimated model is of the form:

$$Erosion_{ijtp} = \beta_0 + \beta_1 * \text{Pioneer legal activity}_{itp} + \beta_2 * \text{Follower legal activity}_{jtp} + \beta_3 * \text{Months in monopoly}_p + \beta_4 * \text{Order}_{jp} + \beta_5 * \text{Penetration}_{tp} + \beta_6 * \text{Growth}_{tp} + \beta_7 * \text{Market share of pioneer}_{i(t-1)p} + \varepsilon_{ijtp}$$

The analysis of longitudinal data with ordinary least squares is subject to violations of the conventional suppositions, mainly homocedasticity and auto-correlation. Not taking account of these problems may cause the estimations to be biased, resulting in an inflated statistical F value (Bergh and Holbein, 1997). Working with information on firms from different countries–as here–makes a problem of heterocedasticity highly likely. The Cook-Weisberg statistic indicated the presence of heterocedasticity when using models of ordinary least squares. The Wooldridge test confirmed the existence of an auto-correlation problem in the data. Crossed-sectional time-series linear models using feasible generalized least squares were used to get around both problems. After calculating the variance inflation factors (VIF) for all the variables, none of them was found to be higher than 2. This indicates that multicollinearity is not a problem in this study.

The results obtained for both pioneers and followers in the European mobile telecommunications industry (see table 5) show the effectiveness of legal actions for attacking rivals and improving competitive positions.

The data show that pioneers that have filed more lawsuits succeeded in sustaining their competitive advantage–measured in terms of market share–for a longer period of time. The study also reveals, however, that the more litigious followers managed to erode pioneer advantage more intensely. Lastly, the results of the regression analysis confirm that the time in monopoly helps preserve the first-mover advantage.

To sum up, this research confirms the effectiveness of legal actions in the fledgling European

Table 5. Regression results[a,b]

	Model 1	
Variable	β	*Std. Dev.*
Pioneer legal activity	0.094	0.025***
Follower legal activity	-0.038	0.019**
Months in monopoly	-0.004	0.001***
Order	0.158	0.035***
Penetration	0.001	0.001
Growth	0.002	0.001
Market share of pioneer "i"$_{t-1}$	0.539	0.167***
Constant	-0.459	0.125***
Log likelihood	-31.710	
X^2 (d,f,)	57.50(7)***	

[a] *N=183 in all models.*
[b] **p<0.01; **p<0.05; ***p<0.01.*

mobile telecommunications industry. Despite divergent interests, these types of NMA help both pioneers and followers reach their objectives.

FUTURE TRENDS AND CONCLUSIONS

This chapter has underlined the important role that non-market actions (NMAs) play in building competitive advantages in regulated industries. The interaction between market and non-market environments forces firms to act on both fronts. Thus, firms looking to develop long-term advantages and to improve performance should undertake market actions such as launching new products, marketing campaigns, price cutting, or introducing new services. In addition, however, firms can also turn to NMAs such as lobbying or legal actions. The use of this type of action is becoming ever more common as a result of the growing pressure stakeholders exert on firms. Government and regulatory agencies mark out the competitive ground rules, but private activists and NGOs are now beginning to play a determining role in many industries, which is forcing firms to widen their range of actions.

Firms need to improve their knowledge of how to design and implement non-market strategies. To do this, they must study the relationships between different NMAs and competitive advantages. The most important types have been identified: information, financial contributions, constituency building and legal actions. Each action has a different purpose, which is why choosing the most appropriate action is crucial for creating competitive advantages. Legal actions in particular work to hinder the entry of competitors into the industry and moderate the effect of regulation on firms, while they can also be used to reduce market power.

The mobile telecommunications industry is a regulated fledgling industry, with sustainable first-mover advantages. This situation makes NMAs particularly effective competition tools in this industry. Pioneer operators use them to protect themselves from new competition, while followers have no choice but to employ the same actions to either improve or create an advantageous competitive position. Specifically, legal actions have been shown to have a statistically significant impact on the competitive position of both pioneers and followers. Pioneer operators undertaking a greater number of legal actions manage to sustain their competitive advantage, while follower firms that more frequently complain to regulatory bodies and courts manage to erode market share from pioneers.

The practical implications of this study are clear. Although litigation is costly for firms, it is an effective competitive weapon. In addition to legal action, the erosion of pioneer competitive advantage depends on the time in monopoly. Therefore, governments should open markets and eliminate all barriers to competition as quickly as possible.

This chapter has focused on the use of legal action within the mobile telecommunications industry. Future studies should investigate the possible generalization of these results to other industries. Extending the study to other types of NMA, along with their implementation and possible applications in the mobile telecommunications industry, would also be highly interesting. Likewise, a study on how changes in regulation affect the use of NMAs would be useful. In summary, a broader strategic analysis that includes other forms of competing is needed.

REFERENCES

Baron, D. P. (1993). *Business and its environment.* Englewood Cliffs, NJ: Prentice Hall.

Baron, D. P. (1995). Integrated strategy: Market and nonmarket components. *California Management Review, 37*, 73-85.

Bhagat, S., Bizjak, J., & Coles, J. L. (1998). The shareholder wealth implications of corporate lawsuits. *Financial Management, 27,* 5-27.

Bharadwaj, S., Varadarajan. G. R., & Fahy, J. (1993). Sustainable competitive advantage in service industries: A conceptual model and research propositions. *Journal of Marketing, 57*(4), 83-99.

Bergh, D. D., & Holbein, G. F. (1997). Assessment and redirection of longitudinal analysis: Demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal, 18*(7), 557-571.

Bizjak, J. M., & Coles, J. L. (1995). The Effect of Private Antitrust Litigation on the Stock-Market Valuation of the Firm. *The American Economic Review, 85*, 436-461.

Bonardi, J. P. (1999). Market and nonmarket strategies of a former monopoly during deregulation: The British Telecom Case. *Business and Politics, 1*, 203-232.

Capron, L., & Chatain, O. (2008). Competitors´resource oriented strategies: Acting on competitors´resources through interventions in factor markets and political markets. *Academy of Management Review, 33*(1), 97-121.

De Figueirido, R. J. P., & Edwards, G. (2007). Does Private Money Buy Public Policy? Campaign Contributions and Regulatory Outcomes in Telecommunications. *Journal of Economics and Management Strategy, 16*(3), 547-576.

El País (2006, November 20). *Las compañías telefónicas pierden siete de cada ocho pleitos judiciales contra el organismo regulador,* Madrid.

European Union (1994). *Towards the Personal Communications Environment: Green Paper on a common approach to mobile and personal communications in the European Union*: COM (94) 145 final. Brussels, Commission of the European Communities.

European Union (2002). *On a common regulatory framework for electronic communications networks and service. Directive Framework* 2002/21/EC. Brussels, Commission of the European Communities.

European Union (2007). *European Electronic Communications Regulation and Markets 2006 (12th report).* (SEC[2007]403), Brussels, Commission of the European Communities.

Fernández, Z., & Usero, B. (2007). The erosion of pioneer advantage in the European mobile telecommunications industry. *Service Business, 1*(3), 195-210.

Field, L., Lowry, M., & Shu, S. (2005). Does disclosure deter or trigger litigation? *Journal of Accounting and Economics, 39*, 487-507.

Frynas, E. G., Mellahi, K., & Pigman, G. A. (2006). First mover advantages in International Business and Firm-specific political resources. *Strategic Management Journal, 27,* 321–345.

Hillman, A. J., & Hitt, M. A. (1999). Corporate Political Strategy Formulation: A Model of Approach, Participation, and Strategy Decisions. *The Academy of Management Review, 24*(4), 825-842.

Jauch, L., Osborn, R., & Martin, T. (1980). Structured content analysis of cases: A complementary method of organizational research. *Academy of Management Review, 5*, 517-526.

Karpoff, J. M., & Lott, J. R. (1999). On the determinants and importance of punitive damage awards. *Journal of Law and Economics, 42*, 527-573.

Koku, P. S., Qureshi, A. A., & Akhigbe, A. (2001). The effects of news on initial corporate lawsuits. *Journal of Business Research, 53*, 49-55.

Laffont, J. J., Rey, P., & Tirole, J. (1997). Competition between telecommunications operators. *European Economic Review, 41*, 701-711.

Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. *Strategic Management Journal, 9,* 41-58.

Schuler, D., Rehbein, K., & Cramer, R. (2002). Pursuing strategic advantage through political means: A multivariate approach. *Academy of Management Journal, 45,* 659–672.

Shell, G. R. (2004). Make the Rules Or Your Rivals Will. Crown Business: New York.

Shell, G. R., & Yao, D. (2000). *Strategy in a World with Uncertain Rules*, working paper. Philadelphia, University of Pennsylvania: The Wharton School.

Smith, K. G., Grimm, C. M., & Gannon, M. J. (1992). *Dynamics of competitive strategy.* London: Sage Publications.

Usero, B., & Fernández, Z. (2005). *First come, first served: An analysis of pioneer and follower firms′s market and non-market actions in the European mobile telephone industry* (Business Economics 05-48-12). Madrid, Spain: Universidad Carlos III de Madrid.

Yoffie, D.B., & Bergenstein, S. (1985). Corporate political entrepreneurship. *California Management Review, 28,* 1, 313-337.

KEY TERMS

First-Mover Advantages: Benefits obtained by firms for being the first firms to enter into an industry. Their primary sources are pre-emption of assets, technological leadership and buyer switching costs.

Non-Market Actions: Measures adopted by a firm in a non-market environment (e.g., legal actions, testifying at government hearings, lobbying, advocacy advertising, or constituency building).

Non-Market Actions — Constituency Building: Informing and encouraging those who could be affected by public policies that have an impact on the firm to embark on political action.

Non-Market Environment: Set of political, social and legal agents interacting with the firm outside of, but in conjunction with, the markets.

Non-Market Actions — Financial Contributions: Providing financial support for politicians. They include measures such as funding political parties or hiring employees with political experience.

Non-Market Actions — Information: Providing information to decision makers (politicians, regulators). They include measures such as lobbying, reporting research and survey results, testifying as experts and supplying technical reports.

Non-Market Actions — Legal Actions: Formal complaints concerning allegedly infringed rights or dissatisfaction with the existing legislation. They include measures such as filing lawsuits against rivals or regulatory authorities, and regulatory petitions.

APPENDIX A: OPERATORS IN THE EUROPEAN MOBILE TELECOMMUNICATIONS INDUSTRY (2008)

Country	Operator*	Launch
Austria	A1 (Mobilkom Austria)*	December 93
	T Mobile Austria	July 96
	One	October 98
	Tele.ring (acquired by Mobilkom Austria)	May 00
	3 AT	May 03
Belgium	Proximus*	January 94
	Mobistar (Orange)	August 96
	Base	March 99
Denmark	TDC Mobil*	July 92
	Sonofon (Telenor)	July 92
	Telia DK	June 97
	3 DK	October 03
Finland	Elisa	December 91
	Sonera*	June 92
	Alands Mobiltelefon	February 93
	DNA	January 01
France	Orange France*	July 92
	SFR	December 92
	Bouygues Telecom	May 96
Germany	Vodafone	June 92
	T-Mobile*	July 92
	E-Plus	May 94
	O_2	October 98
Greece	Vodafone	July 93
	Wind	July 93
	Cosmote*	June 98
	Q-Telecom	June 02
Ireland	Vodafone*	July 93
	O2	March 97
	Meteor	February 01
	3	July 05
Italy	TIM*	April 95
	Vodafone	September 95
	Wind	March 99
	3	March 03

Country	Operator	Launch
Luxemburg	LuxGSM (P&T) *	July 93
	Tango	May 98
	Vox Mobile	April 05
Netherlands	KPN*	July 94
	Vodafone	September 95
	T-Mobile (including Orange)	February 99
Norway	Telenor*	May 93
	NetCom	September 93
	Network Norway	May 06
Portugal	TMN*	October 92
	Vodafone	October 92
	Optimus	August 98
Spain	Movistar*	July 95
	Vodafone	October 95
	Orange	January 99
	Yoigo	December 06
Sweden	Telenor	September 92
	Tele 2	September 92
	Telia*	November 92
	3	August 02
	Swefour	February 04
Switzerland	Swisscom*	March 93
	Orange	June 99
	Sunrise	December 98
	In&Phone	June 05
United Kingdom	Vodafone	July 92
	T-Mobile	September 93
	Orange	April 94
	O2*	December 94
	3	March 03
	MCom	September 06
	PMN	September 06
	Opal Telecom	November 06

**Subsidiaries of the former monopolies in fixed telephony (source: GSM Association and the trade journal Mobile Communications).*

APPENDIX B: COLLECTION AND CODING OF DATA

1. **Definition of 'legal actions:'** A group of professionals and academics defined 'legal actions' as, "Formal complaints concerning allegedly infringed rights or dissatisfaction with the existing legislation. They include measures such as filing lawsuits against rivals or regulatory authorities and regulatory petitions."
2. **Identification of news reports that possibly refer to legal actions:** A search of all the news items that may contain information on legal actions was performed (in total 1,510 new reports).
3. **Determination of the reliability of the data:** Three coders read 10 percent of all the news items and coded those that referred to legal actions. This was done to check that the results were identical independently of the coder.
4. **Coding:** The coders read the news items under study and sorted them according to whether they included information on legal actions or not.

APPENDIX C: MEASUREMENT OF VARIABLES

Dependent Variable	
Erosion$_{ijtp}$	This variable was constructed using the pioneer "i" and follower "j" firms' market shares in one country "p" during two consecutive periods of time (Ferrier et al., 1999).
Independent variables	
Pioneer legal activity$_{itp}$	Total number of legal actions initiated by the pioneer "i" in a specific period "t" and country "p".
Follower legal activity$_{jtp}$	Total number of legal actions initiated by the follower "j" in a specific period "t" and country "p".
Control Variable	
Months in monopoly$_{p}$	Number of months the government took to grant a second license in country "p".
Order$_{jp}$	Position in which a follower firm "j" entered the country "p".
Penetration$_{tp}$	Measured as the percentage of potential users in a country "c"who already have a mobile telephone.
Growth$_{tp}$	Change in the number of customers in two consecutive periods in a country "p".
Market share of pioneer$_{i(t-1)p}$	Market share of the pioneer "i" in the period t-1 in country "p".

Chapter XIII
Spatiality and Political Economy of the Global Fiber Optics Industry

Barney Warf
University of Kansas, USA

ABSTRACT

Fiber optics forms the pivotal telecommunications technology of the contemporary global economy, offering greater speed and security than other modes. This chapter reviews the historical development of this communications technique. It then views its rapid growth within the context of contemporary globalization. Fiber's role in contemporary urban restructuring is noted. The chapter then turns to the spatial distribution of the world's fiber lines, noting major transatlantic and transpacific markets and newer systems. The enormous construction boom of the 1990s and early 2000s, however, led to severe overcapacity, with significant economic fallout.

INTRODUCTION

The core of the global telecommunications infrastructure is an extensive and seamlessly integrated network of fiber optics lines. Indeed, far more than any other technology, such as copper cables, microwaves, or satellites, fiber optics supply the vast bulk of data, voice, and video transmission services around the world. Because of their capacity to deliver high volumes of information rapidly and securely (e.g., via broadband), fiber optics cables form the backbone of the Internet as well as private corporate lines, and are widely used in the electronic media for commercial and residential purposes (e.g., cable television). The technology is thus central to understanding contemporary economic, political and cultural transformations.

The objective of this chapter is to offer a reasonably comprehensive overview of fiber optics as a technology, an industry, and a force within the contemporary global economy. It begins with a brief history of how this phenomenon came to be, including the long history of scientific innovation behind it. Second, it situates and contextualizes fiber optics within the contemporary information-intensive global economy. Unfortunately, this issue has often been approached in apolitical and technocratic terms that ignore the social origins and consequences of the industry. Third, it turns briefly to the urban dimensions of this technology, the ways in which it is implicit in folding and refolding the spatiality of urban accessibility. Fourth, it maps out the global geography of fiber optics, focusing on the two major markets across the Atlantic and Pacific Oceans. Fifth, it explores three consequences of the fiber boom of the 1990s, including a wave of corporate failures, the emergence of dark fiber, and the challenge to the satellite industry.

A BRIEF HISTORICAL OVERVIEW OF FIBER OPTICS

Fiber optics are long, thin, flexible, highly transparent rods of quartz glass (or less commonly, plastic) about the thickness of a human hair that can transmit light signals through a process of internal reflection, which retains light in the core and transforms the cable into a waveguide (Agrawal 2002; Freeman 2002; Crisp and Elliot 2005). They can transmit voice, video, or data traffic at the speed of light (299,792 km/sec.); because light oscillates much more rapidly than other wavelengths (200 trillion times per second in fiber cables v. two billion per second in a cellular phone), such lines can carry much more information than other types of telecommunications. Modern fiber cables contain up to 1,000 fibers each and are ideal for high-capacity, point-to-point transmissions. Moreover, fiber cables do not corrode or conduct electricity, which renders them immune to electromagnetic disturbances such as thunderstorms.

Their development reflects a long history of experimentation and technological change. The origins of fiber optics go back to Jean-Daniel Colladon at the University of Geneva, who demonstrated light guiding in 1841. Subsequent experiments in 1870 by British physicist John Tyndall, who used moving water through curved rods to conduct light, showed that optical signals could be bent and that light therefore did not need always to travel in a straight line. In 1880, William Wheeling patented the method of "piping light" through mirrored pipes. Alexander Graham Bell's "photophone" in the 1880s transmitted voice signals on a beam of light; the concurrent introduction of Thomas Edison's light bulb enhanced the popularity of technologies of light. In the 1920s, Scottish television inventor John Baird and Clarence Hansell in the U.S. patented the idea of using transparent rods to transmit images (Hecht 1999). In the 1950s, experiments by Brian O'Brien at the American Optical Company and Narinder Kapany (who coined the term "fiber optics") at the Imperial College of Science and Technology in London developed a fiberscope, or forerunner to contemporary fiber optics, a technology that led to laparoscopic surgery. The introduction of a dense coat, or cladding, around the glass core, by Lawrence Curtiss of the University of Michigan, prevented the loss of light and led to near-perfect internal reflection within the core of the cable. In the 1960s, the use of laser diodes in helium-neon gas perfected this technique at Bell Labs in New Jersey. In 1956 British physicist Charles Kao showed that light attenuation was caused by impurities in the glass and suggested optimal maximum levels of glass purity for long distance transmission. Ten years later, Robert Maurer, Donald Keck and Peter Schultz of the Corning Glass Works (later Corning, Inc., now the largest provider of fiber cable in the world) developed rods of pure fused silica that greatly reduced light attenuation to the levels that Kao specified. In 1960, Theodore Maiman of the Hughes Research

Laboratories in Malibu, California produced the first operational laser.

As computer equipment became rapidly more sophisticated and widespread, U.S. military uses of fiber optics began as it deployed them for communications and tactical systems. In 1975 computers at the NORAD headquarters in Cheyenne Mountain were linked by fiber optics. The technology was also central to the development of the Internet. Indeed, much of the durability and reliability of the Internet reflects its military origins, for its original purpose was to allow communication among computers in the event of nuclear war.

Simultaneously, the microelectronics revolution initiated enormous decreases in the cost of computers and exponential increases in their power and memory, making communications the primary bottleneck to corporate productivity. As fiber optics increasingly appeared to meet rising demand in this sector, corporate applications rose steadily (Olley and Pakes 1996; Jorgenson 2001). In 1977, AT&T installed the first telephone lines to use fiber optic cables, a network 25 miles in length that could carry 672 voice channels, beneath downtown Chicago; GTE followed immediately in Boston (Goff 2002). However, it was during the massive global changes in the world economy at the end of the 20th century that fiber came into its own as the dominant medium of telecommunications. Telephone companies and other providers of telecommunications services began rapidly replacing older copper wire cables with fiber optics, which many observers expect will become virtually the only telecommunications transmission technology in the future. Fiber optics facilitated the explosive growth of e-commerce, which includes both business-to-business transactions as well as those linking firms to their customers, including electronic data interchange (EDI) systems, digital advertising, online product catalogues, the sharing of sales and inventory data, submissions of purchase orders, contracts, invoices, payments, delivery schedules, product updates, and labor recruitment. Indeed, fiber optics arguably transformed the Internet from a communications to a commercial system, accelerating the pace of customer orders, procurement, production, and product delivery (Malecki 2002). In addition, fiber optics are used in a variety of scientific and medical equipment.

Fiber cable itself comprises are relatively small share of the total cost of an undersea cable system. Thus, improvements in fiber optics capacity and efficiency in the 1990s rested on other components of the system. Because signals inescapably attenuate during transmission, repeaters are necessary to maintain the fidelity of optical signals. The first generation of repeaters converted optical signals into electronic voltage in order to amplify them, then reconverted them to optical signals; early fiber cables required frequent repeaters, often every 5 to 10 km. As the purity of fiber cables improved, and as repeaters improved in power, the need for repeaters decreased accordingly. In 1991, optical amplifiers, which remove the need to convert light to electronic signals, such as the erdium doped fiber amplifier (EDFA), improved the efficiency of transmission over electronic amplifiers by a factor of 100. The TAT-12 line, installed in 1995, was the first long-haul system to use EDFA technology. Today, in long-haul cables, repeater distances range as high as 500 to 800 km. Similarly, dense wavelength division multiplexing (DWDM), first developed in the 1970s, made it possible to transmit multiple wavelengths over a single fiber. As a result of these numerous improvements, fiber's bandwidth capacity increased more than 200-fold, from 10 mbps in the 1970s to as high as 50 terahertz per second (thzps) in 2005.

THEORIZING THE GROWTH OF FIBER-BASED CAPITALISM

As numerous observers have pointed out, global capitalism in the late 20th century underwent an enormous sea-change. Telecommunications constitute an integral part of this transformation. The ability to transmit vast quantities of information in real time over the planet is crucial to what Schiller (1999) calls digital capitalism. No large corporation could operate today in multiple national markets simultaneously, coordinating the activities of thousands of employees within highly specialized corporate divisions of labor, without access to sophisticated channels of communications. The exploding demand for high bandwidth communications has been a major force behind the growth of the international communications infrastructure. For Castells (1996), this transformation is mirrored in the space of flows and the new geometries that accompany it, which wrap places into highly unevenly connected networks, typically benefiting the wealthy at the expense of marginalized social groups. Ruggie (1993:141) likens such networks to the "economic equivalent of relativity theory." However, the global space of flows is far from randomly distributed over the earth's surface: rather, it reflects and reinforces existing geographies of power concentrated within specific nodes and places, such as global cities, trade centers, financial hubs, and corporate headquarters. Indeed, because the implementation of fiber lines reflects the powerful vested interests of international capital, these systems may be seen as "power-geometries" (Massey 1993) that ground the space of flows within concrete historical and spatial contexts.

Financial and producer services firms were at the forefront of the construction of fiber networks in large part because they allowed the deployment of electronic funds transfer systems, which comprise the nervous system of the international financial economy, allowing banks to move capital around a moment's notice, arbitrage interest rate differentials, take advantage of favorable exchange rates, and avoid political unrest (Langdale 1989; Warf 1995). Fiber carriers are heavily favored by large corporations for data transmissions and by financial institutions for electronic funds transfer systems, in large part because of the higher degrees of security and redundancy this medium offers (Nijkamp and Vleugel 1993). Such networks give banks an ability to move money around the globe at stupendous rates: subject to the process of digitization, information and capital become two sides of the same coin. Liberated from gold, traveling at the speed of light, as nothing but digital assemblages of zeros and ones, global money performs a syncopated electronic dance around the world's neural networks in astonishing volumes. In this context, finance capital is not simply mobile, it is *hypermobile*, i.e., it moves in a continual surge of speculative investment that never materializes in physical, tangible goods. The world's currency markets, for example, trade more than $1 trillion every day, dwarfing the $25 billion that changes hands daily to cover global trade in goods and services. In the securities markets, fiber optics facilitated the emergence of 24 hour/day trading, linking stock markets through computerized trading programs.

Deregulation was also a fundamental part of the growth of the global fiber optics system. This process was initiated by the U.S. with the breakup of AT&T in 1984, which had long enjoyed a monopoly over domestic telephony and was broken up by an antitrust suit. Deregulation opened the door for a proliferation of new fiber optics service providers such as MCI, which grew to become the second largest provider in the world. Sprint arose as the first corporate telecommunications provider entirely based on fiber optics; others such as Qwest followed shortly (Figure 1). In the U.S., the 1996 Telecommunications Act further eliminated regulatory oversight, effectively ending the boundaries between local and long distance traffic and opening the door to a wave of mergers and acquisitions (Warf 2003). Soon thereafter, British Telecommunications, France Telecom, and Deutsche Telekom were partially or totally sold *en masse* to private investors, and in Japan, the monopoly long held by Nippon Telegraph and Telephone (NTT) was broken by government fiat

Figure 1: Qwest National Fiber Network (source: www.alliancedatacom.com)

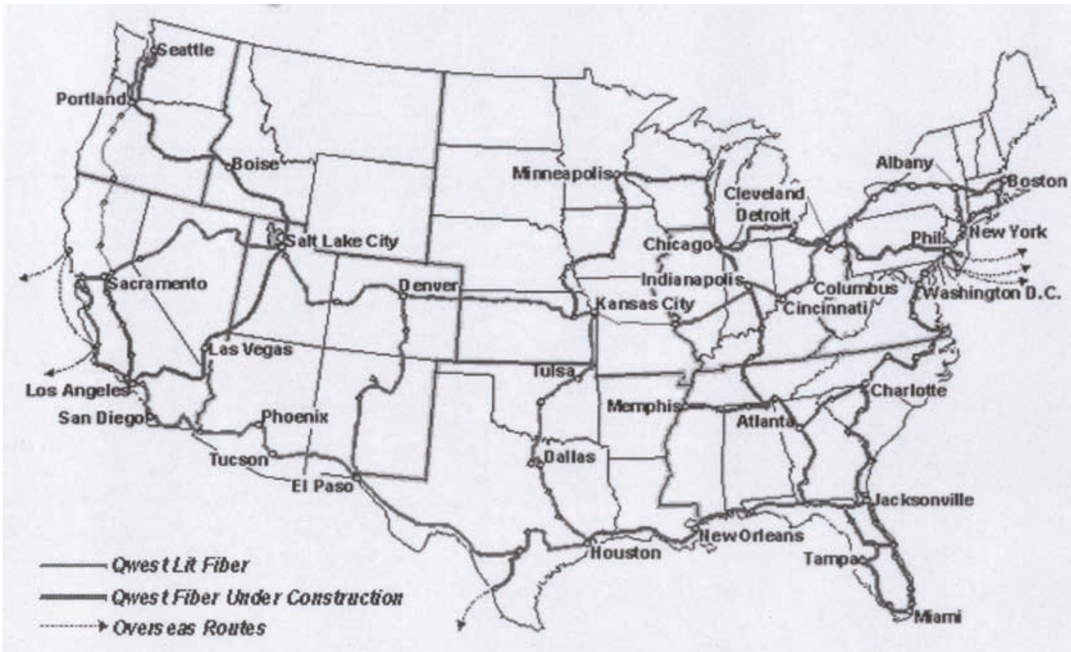


Table 1. Major Trans-Atlantic and Trans-Pacific fiber optics cables

Name	Operational Capacity	Date	Landing Station Locations
TransAtlantic:			
TAT-8	560 mbps	1988	U.S., U.K., France
PTAT-1	1.26 gbps	1989	U.S., U.K., Bermuda, Ireland
PTAT-2	1.26 gbps	1992	U.S., U.K.
TAT-9	1.12 gbps	1992	U.S., U.K., France, Spain, Canada
TAT-10	1.12 gbps	1992	U.S., Germany, Netherlands
TAT-11	1.12 gbps	1993	U.S., U.K., France
TAT-12	5 gbps	1995	U.S., U.K.
TAT-13	5 gbps	1995	U.S., France
Gemini	2.5 gbps	1998	U.S., U.K.
AC-1	2.5 gbps	1999	U.S., U.K., Germany
Columbus 3	2.5 gbps	1999	U.S., Spain, Portugal, Italy
TAT-14	10 gbps	2000	U.S., U.K., France, Netherlands, Germany
FLAG Atlantic	10 gbps	2001	U.S., U.K., France
Apollo	10 gbps	2002	U.S., U.K., France
TransPacific:			
HAW-4/TPC-3	560 mbps	1989	California, Hawaii, Guam, Japan
GPT	280 mbps	1989	Guam, Philippines, Taiwan
H-J-K	280 mbps	1990	Hong Kong, Japan, S. Korea
NPC	1.26 gbps	1990	Oregon, Alaska, Japan
TASMAN-2	1.12 gbps	1991	Australia, New Zealand
TPC-4	1.12 gbps	1992	California, Canada, Japan
HAW-5	1.12 gbps	1993	California, Hawaii
PacRim East	1.12 gbps	1993	Hawaii, New Zealand
PacRim West	1.12 gbps	1994	Australia, Guam
TPC 5/6	5 gbps	1995	California, Oregon, Hawaii, Japan
KJG	1.12 gbps	1995	S. Korea, Japan, Guam
TPC-5	5 gbps	1996	California, Hawaii, Guam, Japan
Southern Cross	2.5 gbps	1999	California, Hawaii, Fiji, Australia
China-US	2.5 gbps	1999	California, Hawaii, Guam, S. Korea, Japan, China, Taiwan
PC-1	10 gbps	2000	Japan, U.S.
Japan-US	10 gbps	2000	Japan, U.S.
FLAG Pacific 1	10 gbps	2002	Japan, U.S., Canada

Source: Smith 2003.

(although like France Telecom, it remains largely publicly owned). The World Trade Organization's Basic Telecommunications Agreement, which went into effect in 1998, also fostered competition around the world. Today, state-owned or regulated telecommunications monopolies are increasingly rare around the world. In 2005, roughly 1,000 fiber optics and two dozen public and private satellite firms competed to provide international telecommunications service, the vast majority of which originated in economically developed countries. The consequences for the market structure of telecommunications were dramatic, including new competitors, improved service, and rapidly falling costs, although Graham and Marvin (1996) note that in this climate, providers may freely engage in "cherry picking," i.e., servicing only high-profit clients at the expense of the needy and disempowered.

Large fiber networks are generally owned and operated by consortia of firms. Until the 1990s, all commercial fiber lines were built, used, and paid for by a handful of monopoly carriers such as AT&T, British Telecom, Japan's Kokusai Denshin Denwa (KDD), known informally in the industry as "The Club." The Club system allowed telecommunications carriers to construct and own undersea cables and to serve as their users or vendors. Typically, landing facilities are owned by carriers from the country in which the facility is located but the "wet links" (undersea cables) are jointly owned by club members. Under the club system, AT&T, for example, ventured aggressively into the international fiber optics market as it globalized in the face of declining market share in the U.S., often by entering strategic alliances that stretched across national borders (Warf 1998). Similarly, Sprint affiliated with France Telecom and Deutsche Telekom to form Global One in 1996, and AT&T and British Telecom acquired a 30% share of Japan Telecom. Table 1 lists the major submarine cable networks in place in 2003 for the two largest markets across the Atlantic and Pacific Oceans. Under the Club system, capacity was allocated and payments made before or during construction of the network. Members were required by national regulators to sell capacity to non-members on a non-discriminatory basis close to cost. Allegations arose that Club members discriminated against new entrants by offering disadvantageous conditions of membership, such as capacity prices. However, as deregulation encouraged new entrants into the cable markets, the Club system began to fragment. Private systems, in which carriers invite non-carrier investors such as banks, emerged as an alternative system, and recently, non-carrier systems have also appeared.

URBAN GEOGRAPHIES OF FIBER OPTICS

Starting in the 1980s, telecommunications firms began to build a large interurban network of fiber optics lines in the U.S., whose aggregate networks exceeded 50 million km by 2001. The largest fiber optic lines (T3, OC-3, OC-4, and OC-12) lines connect a handful of large metropolitan areas, whose comparative advantage in producer services has benefited significantly by publicly-installed telecommunications systems. (For examples, see http://cybergeography.planetmirror.com/cables.html). While the largest metropolitan regions are well served (particularly New York, Chicago, Washington, DC, Atlanta, Los Angeles, and Seattle), many other areas (such as the rural South) have few connections. High capacity fiber lines are particularly important in regard to access to high-density material, e.g., graphical content on the WWW. For high volume users (typically large service firms), for whom the copper cables used by telephone companies are hopelessly archaic, these lines are an absolute necessity. For large real estate developers, fiber capability has emerged as a critical issue in determining the price and attractiveness of corporate office space, indicating that relative space via connectivity is as important as accessibility via conventional transportation. Moreover, numerous cities have taken the initiative to establish their own municipal fiber networks as part of their economic development strategies to attract firms rather than wait for the private sector, often in the form of public-private partnerships. In such cases, fiber lines are often packaged along with the other municipal utilities such as water, electricity, or natural gas. Thus, a grid of fiber lines surrounding the core of cities has become an indispensable part of urban comparative advantage. Rural areas, in contrast, often suffer a distinct disadvantage in terms of this digital divide (Gabe and Abel 2002).

Fiber optics providers prefer large metropolitan regions where dense concentrations of corporate and residential clients allow them to realize significant economies of scale and where frequency transmission congestion often plagues satellite traffic. So-called “global cities” such as New York, London and Tokyo (Sassen 1991) are prime beneficiaries, using fiber optics lines to spread their sphere of influence around the planet. For example, the Atlanta metropolitan region exhibits 400,000 miles of fiber optic lines, which have been important to the revival of downtown regions and enhanced its competitive position within the national urban hierarchy (Walcott and Wheeler 2001). Within cities, fiber lines accelerate the creation of wealth by corporate elites, generating geographies of inequality in which the wired and the wireless, the haves and have-nots of the information, live in close proximity; even in the most networked of cities, there exist large disenfranchised groups who pay the costs of the digital economy but reap relatively few of the benefits. In contrast with metropolitan areas, rural areas, with relatively small populations and low market potential, hold little market appeal. This urban bias, and the social schisms it deepens, is replicated at the international scale; Graham (1999) notes that the skein of fiber cables linking the world’s major cities is vital to their role and domination over the world economy. Despite the mythologized notion that fiber optics lines erase spatiality, therefore, it is evident that the geographic impacts of this technology are highly selective.

The growth of fiber optics for commercial and residential purposes, such as cable television, assumes that local lines are effectively linked to high-capacity backbone routes. However, this connection often confronts the “last mile” problem, the gap between a facility or client and a Point of Presence (POP), the point at which the facilities of an inter-exchange carrier are accessible. Telecommunications and cable television companies have devoted substantial resources to overcoming this problem, and as a result, broadband access has improved gradually. Some, such as Verizon, have pioneered the development of fiber-to-the-premises (FTTP) networks.

GEOGRAPHIES OF GLOBAL FIBER OPTICS NETWORKS

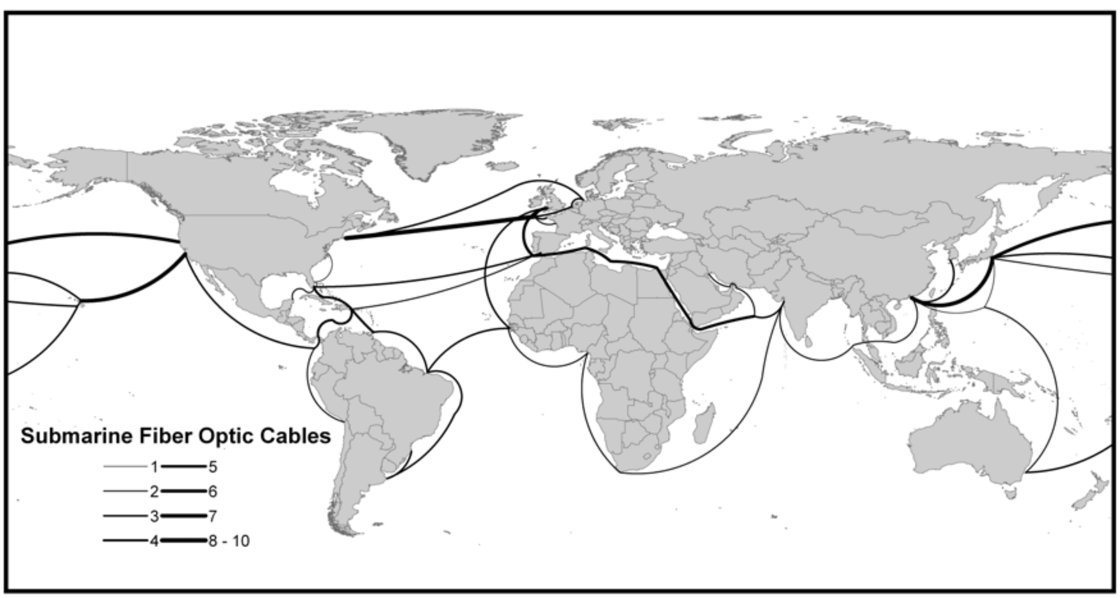
Despite exaggerated claims that telecommunications render distance meaningless (e.g., Cairncross 1997), the geography of fiber optic lines reflects the accumulated imprints of successive rounds of investments in space and time. The placement of terrestrial networks reflects the complex ways in which space, the global economy, and technology are wrapped up in each other. Spurred by the growth of information-intensive services and predictions of unending growth in Internet traffic, telecommunications companies undertook an orgy of fiber optic cable construction in the 1980s and 1990s.

Laying transoceanic fiber cables entailed a host of technical and organizational issues. In addition to the costs of purchasing fiber, telecommunications companies must pay for the laying of fiber across the ocean floor and the installation of “manholes,” on-shore bunkers designed to allow access for repairs. AT&T’s Submarine Systems, the world’s largest supplier of undersea telecommunications systems, operates a fleet of six cable ships to service its 230,000 kilometers of undersea cable. Submarine lines must be routed to avoid seismic activity (earthquakes and undersea avalanches), ships' anchors, deep sea currents, fishing trawlers, and military activities, and must be armored against sharks, which are attracted by electromagnetic emissions. While the original lines were point-to-point, the development of submarine branching units (SBUs) allowed multiple points to be served simultaneously, leading to more complex network configurations. Moreover, most submarine cables today are "self-healing," meaning that they offer redundant capacity and high resiliency, so that the loss of one link can be easily and rapidly compensated by others. Today, the world's fiber system totals more than 25 million km in length, connecting all the world's continents except Antarctica (Figure 2).

The geography of global fiber networks centers primary upon two distinct telecommunications markets crossing the Atlantic and Pacific Oceans, connecting two of the major engines of the world economy, North America and East Asia (Chaffee 2001). In 1988, in conjunction with MCI and British Telecommunications, AT&T initiated the world's first trans-oceanic fiber optic cable, Trans-Atlantic Telecommunications (TAT-8), which could carry 40,000 telephone calls simultaneously. The trans-Atlantic line was the first of a much broader series of globe-girdling fiber lines that AT&T erected in conjunction with a variety of local partners. Because large corporate users are the primary clients of such networks, it is no accident that the original and densest web of fiber lines connects London and New York, a pattern that extends historically to the telegraph and telephone (Hugill 1999). The next generation, TAT-9 and TAT-10, which began in 1992, could carry double the volume of traffic of TAT-8. The third generation, TAT-11 to TAT-13, was the first to use EDFA rather than older repeaters. Newer generations of cable were even more powerful.

Starting with the Trans-Pacific Cable (TPC-3) in 1989 connecting the New York and Tokyo stock exchanges, a growing web of trans-Pacific lines mirrored the rise of East Asian trade with North America, including the surging economies of the Newly Industrialized Countries. In 1996, the first all-fiber cable across the Pacific,

Figure 2. The World's Major Fiber Optic Cables, 2003 (source: after Staples 2004)



TPC-5, was laid. In 2006, a consortium including Verizon and five Asian providers announced plans to lay an 11,000 mile U.S.-China link that would support 1.28 terabits of information – 60 times the capacity of the next largest cable – in time for the Beijing Olympics in 2008 (Shannon 2006). In 2007, Google announced the purchase of large quantities of trans-Pacific fiber cable with the aim of launching a multi-terabit Unity service in 2009.

The complex interplay of deregulation, globalization, and technological change increased the international transmission capacities and traffic volumes for fiber optics carriers explosively (Figure 3). Between 1988 and 2003, for example, trans-Atlantic fiber optic cable capacity increased from 43,750 voice paths to 45.1 billion (103,000 percent), while across the Pacific Ocean, cable carriers' capacity rose from 1,800 voice paths to 1.87 billion (an astonishing 1.6 billion percent).

In addition to the two major markets, fiber lines have extended into several newer ones. In 1997, AT&T, NYNEX and several other firms (including, for the first time, non-telecommuni-

Figure 3. Growth in TransAtlantic and TransPacific Fiber Capacities, 1990-2003 (source: calculated by author from data in Staples 2004)

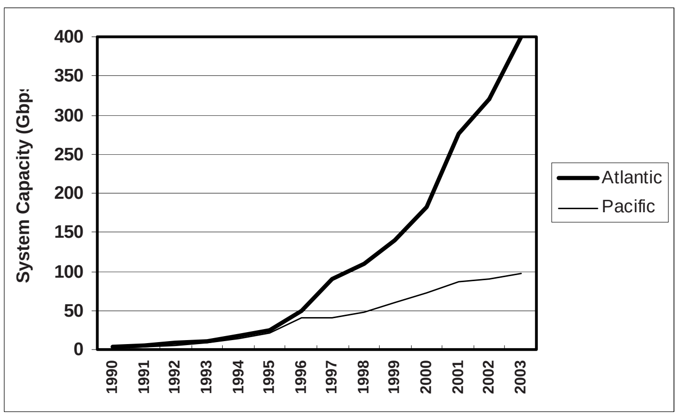


Figure 4. The FLAG Network

Figure 5. The Africa ONE Fiber System (source: http://news.bbc.co.uk)

cations firms) opened the self-healing Fiberoptic Link Around the Globe (FLAG), a system that eventually expanded to 55,000 kilometers connecting Europe and Southeast Asia (Figure 4). The world's longest submarine telecommunications network (Denniston 1998), FLAG, the world's longest submarine telecommunications cable, filled a void in undersea cable capacity between Europe, the Middle East, and Asia. It also hooked into regional systems such as the Asia Pacific Cable Network, a 12,000 kilometer system linking Japan, South Korea, Taiwan, Hong Kong, the Philippines, Thailand, Vietnam, and Indonesia, as well as the Caribbean Fiber System (i.e., the Eastern Caribbean Fiber System, Antillas 1, Americas 1, and Columbus 2). Unlike earlier systems, FLAG allowed carriers to purchase capacity as needed, rather than compelling them to purchase fixed quantities.

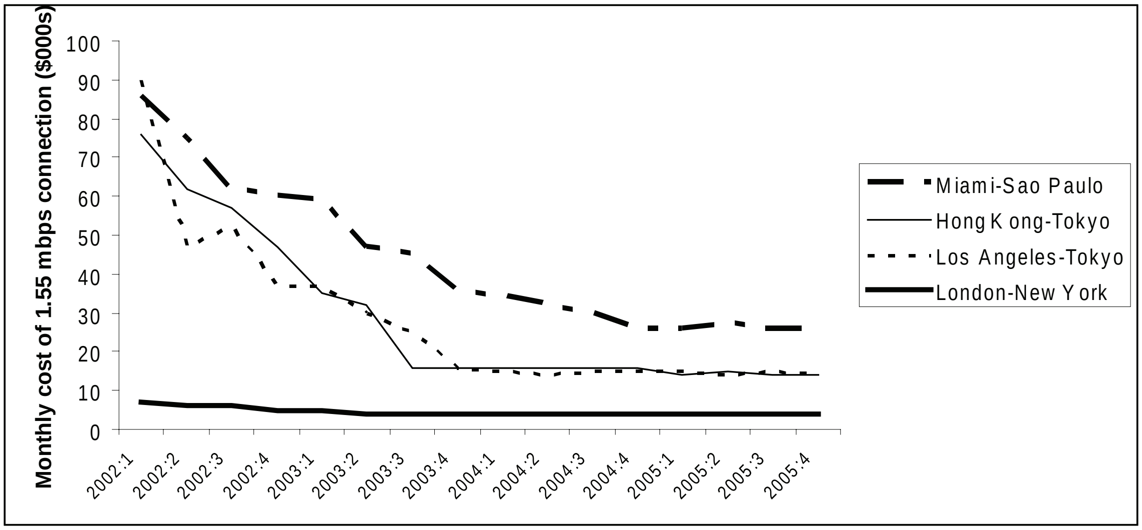
In Africa, long marginalized in global telecommunications, AT&T's $2 billion Africa ONE (*O*ptical *NE*twork) system circumnavigates the continent (Marra and Schesser 1996; Noam 1999) with a 32,000-39,000 km (estimates vary), high-capacity, self-healing undersea fiber optic network (Figure 5). With a capacity of 2.5gbps, Africa ONE is designed to link roughly 41 landing points in 27 coastal countries with large multinational corporate clients internationally. Although the continent as a whole is impoverished, the system's providers sense high potential profits in that specific market. However, the system has met repeated financial difficulties; set to begin in 2002, it had yet to initiate operations in 2005.

THREE CONSEQUENCES OF THE FIBER OPTICS BOOM

The massive surge of supply in the global fiber optics industry generated three distinct, important, but unintended effects: oversupply and economic crisis, the growth of "dark fiber," and a serious challenge to the satellite industry.

As with all industries in which supply increases markedly more rapidly than demand, the explosive growth in the world's fiber optics capacity ultimately led to overcapacity and declining utilization rates (ElBoghdady 2001). As the growth of the world's fiber optics networks' transmission capacity outstripped the rise in demand, transmission prices plunged in a deflationary spiral throughout the first decade of the 21st century, often by as much as 90 percent (Figure 6). Telecommunications corporate stock prices plunged, forcing numerous fiber optics firms into debt and bankruptcy and most others into financial restructuring. The list of casualties from this debacle in the early 2000s included: Global Crossing, Metromedia Fiber Network, Viatel, MCI/Worldcom, Williams Communications, Winstar Communications, and PSINet. Some victims were purchased by buyers eager to become players on the global stage: in 2003, for example, FLAG Telecom was bought by the Indian wireless services provider Reliance. Low fiber transmission prices, in turn, helped to keep down the costs of telephone calls and other applications of the technology.

Figure 6. Declining Prices of Fiber Optics Services, 2002-2006 (source: calculated from data in Telegeography)



Overall capacity utilization rates fell below 50 percent, leading to large quantities of unused "dark fiber." With considerable amounts of dark fiber, corporate clients often lease excess backhaul capacity from former monopolies in order to connect domestic networks to the international system. In addition to system overcapacity, dark fiber reflects the high costs of planning and instilling fiber lines, which leads providers to lay more than necessary in anticipation of rising future demand. For example, a utility company may deliberately install dark fiber in the expectation of leasing it to a cable television company in the future. In addition, however, dark fiber also came to mean the leasing of unused fiber capacity from network service providers. (Indeed, some companies specialize in this market).

One consequence of the explosion of fiber capacity was mounting competition with the besieged satellite industry, with which fiber optics are quasi-substitutable (Pfeifenberger and Houthakker 1998). While satellites are ideal for point-to-area distribution networks common in the mass media, especially in low-density regions, fiber optic lines are preferable for point-to-point communications, especially when security is of great concern. Before the explosive growth in fiber capacity in the 1990s, satellites were traditionally more cost-effective for transmission over longer distances (e.g., more than 500 miles), while fiber optic lines often provided cheaper service for shorter routes (Langdale 1989). The rise of the integrated global fiber network, however, steadily eroded satellites' share of global traffic in data and video transmission services. Despite the pitch by satellite operators that satellites could provide Internet backbone services as a way to bypass terrestrial congestion, fiber remains by far the preferred technology. Satellites simply cannot offer sufficient security or backup capacity to be economically competitive with fiber. In 2003, fiber optics carriers comprised 94.4 percent of worldwide transmission capacity (up from 16 percent in 1988), including 91.3 percent across the Pacific and 95.2 percent across the Atlantic Ocean (Warf 2006).

CONCLUDING THOUGHTS

Fiber optics are one of the transformative innovations to emerge from the microelectronics revolution of the late 20th century. The development of this technology was the culmination of a long history of research by individuals, universities, the military, and corporations, and led to a mode of telecommunications significantly more powerful, secure, and rapid than competing technologies. However, because technology is a social, not simply technical, phenomenon, the dramatic expansion in fiber optics capacity and utilization reflects the historically-specific circumstances of global capitalism as it emerged from the crises of the 1970s and the end of the post-WWII economic boom. Fiber optics were ideal for the information-intensive nature of financial and producer services, particularly when security is of critical importance, and firms in this sector comprised the driving force behind the demand that propelled a vast global network of lines in the 1980s and 1990s. Whereas the two largest markets are those stretched across the Atlantic and Pacific Oceans, newer networks have increased the reach of fiber optics into Latin America and Africa.

At multiple spatial scales, from the urban to the nation to the world, therefore, fiber optics lines realigned the geographies of centrality and peripherality. Far from annihilating space, therefore, the industry reconfigured it. However, the logic that propelled the industry to such prominence also gave rise to the overcapacity and end of the "dot com" boom in the late 1990s and early 2000s, initiating a severe period of corporate retrenchment and restructuring. Unused capacity – dark fiber – appeared in both planned and unplanned forms. The dramatic decline in prices that accompanied this trend posed severe competitive problems for fiber optics providers, including a wave of bankruptcies, but also eroded the market share held by substitutes to fiber such as satellite services.

What does the future hold for this industry? In the short term, the substantial overcapacity in long-distance fiber generated by the boom of the 1990s will be difficult to overcome. Future market potential likely rests in the provision of services to residences (e.g., cable television and high-speed Internet), if the last mile problem may be conquered effectively. The wireless revolution may also pose a competitive challenge to fiber optics. It is evident from these remarks that fiber exhibits the dynamism and fluidity characteristic of the telecommunications sector as a whole.

REFERENCES

Agrawal, G. (2002). *Fiber-optic communication systems.* New York: Wiley.

Cairncross, F. (1997). *The death of distance*. Boston: Harvard Business School Press.

Castells, M. (1996). *The rise of the network society.* Oxford: Blackwell.

Chaffee, C. (2001). *Building the global fiber optics superhighway.* New York: Springer.

Crisp, J., & Elliot, B. (2005). *Introduction to fiber optics*, 3rd edition. London: Newnes.

Denniston, F. (1998). FLAG – Fiber-Optic Link Around the Globe. *Sea Technology,* Feb., 78-83.

ElBoghdady, D. (2001). Fiber-optic firms face issue of overbuilding. *Washington Post,* February 28, 1.

Freeman, R. (2002). *Fiber optics systems for telecommunications.* New York: Wiley.

Gabe, T., & Abel, J. (2002). Deployment of advanced telecommunications structure in rural America: Measuring the rural divide. *American Journal of Agricultural Economics*, *84*, 1246-1252.

Goff, D. (2002). *Fiber optics reference guide*, 3rd edition. Woburn, MA: Focal Press.

Graham, S. (1999). Global grids of glass: On global cities, telecommunications, and planetary urban networks. *Urban Studies, 36*, 929-949.

Graham, S., & Marvin, S. (1996). *Telecommunications and the city: Electronic spaces, urban places*. London: Routledge.

Hecht, J. (1999). *City of light: The story of fiber optics*. Oxford and New York: Oxford University Press.

Jorgenson, D. (2001). Information technology and the U.S. economy. *American Economic Review, 91*, 1-32.

Maclean, G. (1995). Will fiber optics threaten satellite communications? *Space Policy, 11*, 95-99.

Malecki, E. (2002). The economic geography of the Internet's infrastructure. *Economic Geography, 78*, 399-424.

Marra, W., & Schesser, J. (1996). Africa ONE: The Africa optical network. *Communications Magazine*, *34*(2), 50-57.

Massey, D. (1993). Power-geometry and a progressive sense of place. In J. Bird, B. Curtis, T. Putnam, G. Robertson, & L. Tickner (Eds.), *Mapping the futures: Local cultures, global change* (pp. 59-69). London: Routledge.

Noam, E. (1999). *Telecommunications in Africa*. New York: Oxford.

Olley, G., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*, 1263-1297.

Pfeifenberger, J., & Houthakker, H. (1998). Competition to international satellite communications services. *Information Economics and Policy, 10*, 403-430.

Ruggie, J. (1993). Territoriality and beyond: Problematizing modernity in international relations. *International Organisation, 47*, 139-174.

Sassen, S. (1991). *The global city: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Schiller, D. (1999). *Digital capitalism: Networking the global market system*. Cambridge, MA: MIT Press.

Shannon, V. (2006). Group plans to build China-U.S. fiber optic link. *New York Times* December 19. http://www.nytimes.com/2006/12/19/technology/19cable.html?pagewanted=all

Singhi, M., & Long, H. (1998). New undersea cable developments and satellite services: Toward complementary coexistence in the 21st Century. *Proceedings of the 20th Pacific Telecommunications Conference* (pp. 566-569). Honolulu: Pacific Telecommunications Council.

Smith, D. (2003). *Digital transmission systems*. London: Springer.

Staple, G. (2004). *Telegeography 2004: Global telecommunications traffic statistics and commentary*. Washington, DC: Telegeography, Inc.

Walcott, S., & Wheeler, J. (2001). Atlanta in the telecommunications age: The fiber-optic information network. *Urban Geography, 22*, 316-339.

Warf, B. (1998). Reach out and touch someone: AT&T's global operations in the 1990s. *Professional Geographer, 50*, 255-267.

Warf, B. (2003). Mergers and acquisitions in the telecommunications industry. *Growth and Change, 34*, 321-344.

Warf, B. (2006). International competition between satellite and fiber optic carriers: A geographic perspective. *The Professional Geographer, 58*, 1-11.

KEY WORDS

Fiber Optics: Glass wires that transmit information via rapidly oscillation pulses of light over long distances.

Internet: Worldwide interconnected system of computer networks designed to transmit information of various types, including data, voice, and video.

Overcapacity: Excess supply of a good, in this case fiber optic cable transmission ability, relative to demand, resulting in unused or underused portions.

Dark Fiber: Fiber optic cables that are not in use, resulting from either planned or unplanned overcapacity.

Deregulation: Decrease or removal of government controls in an industry, including prices, ownership, and market penetration.

Globalization: Increase in the volume, scope, and rapidity of international transactions.

Repeaters: Devices designed to capture, amplify, and transmit information along fiber optic cables to minimize signal attenuation.

Chapter XIV
Digital Convergence and Home Network Services in Korea:
Part 1 – Recent Progress and Policy Implications[1]

Hyun-Soo Han
Hanyang University, South Korea

Heesang Lee
Sungkyunkwan University, South Korea

Yeong-Wha Sawng
Electronic and Telecommunications Research Institute, South Korea

ABSTRACT

In South Korea, the home network industry has been the object of great expectations among the government and business community. Thus far, however, it has proved unable to fulfill these expectations. In this study, we categorized three classes of typical services and five industrial participants with regard to Korean home network industry. We first reviewed the government plans on home network services in the early stage of evolution. Second, consumer value expectations and current home network service levels were examined to understand the issues regarding weak diffusion of home network services. Next, we investigated business interests of the industrial participants comprising convergent home network services, and explored the causes of conflicts among them. On the base of the analyses, we offered policy suggestions from digital convergence perspective to help removing the obstacles that hinder business cooperation among industrial participants of the home network industry. Managerial implications for the participating firms were discussed in Part II adopting business model framework.

HOME NETWROK SERVICE EVOLUTION

Home network service is a digital convergence service on the base of integrated computer information technologies, telecom services, and home appliances in residential areas. In the early stage of home network evolution, the main issue was the technology itself to enable the networking of the appliances by means of wire or wireless telecom technologies. Recently, the scope of home network services is extended to include more comprehensive industry convergence applications such as home broadcasting & entertainment, and integrated communication & information services (National Information Society Agency, 2003). In our study, we classified the home network services into three types: home control & automation service, home broadcasting & entertainment service, and home communication & information service. The definitions of each home network service and particular applications are illustrated in Table 1.

It was expected that Korean home network industry would develop into a representative Korean industry that should lead the world market as well as usher in the age of the 30,000 dollar national per capita income (Kim, 2005). Based on this forecast, many major companies, to dominate the related industry beyond 2010, heavily invested on the technology development to secure standardization in its inception stage that should lead strong positioning in the home network industry. Industrial participants of home network services were diverse due to inherent nature of convergence from many industrial sectors. The companies participating in home network services were from various industrial sectors such as telecom, broadcasting, home appliances, construction, and digital contents. In early stage of home network service evolution, the home network services had been initiated mainly by the equipment manufacturers including home appliances and the telecom service providers. Equipment manufacturers laid emphasis on digital home appliances, computer servers, telecom gateways, and integrated networking of different devices, whereas the telecom service industry focused on back-bone high-speed Internet infrastructure, information provision, home entertainment, and integrated services.

Each industrial participant had strategic motivations to go into the home network service industry. Construction companies, as an opportunity to exploit their experiences cultivated through constructing digital enhanced buildings, apartments, and condominiums in the late 1990's, provided provisional home network services that focused on home automation through cooperation with ISPs (Internet Service Providers). Electronic home appliance manufacturers intended to over-

Table 1. Types of home network services

Type	Definition	Items
Home control/ automation service	A service that monitors and controls devices like electronic home appliances, sensors, and lighting as well as the physical status of the house, using a fixed or mobile handset.	Remote control, home security, crime and disaster prevention, and energy management, etc.
Home broadcasting/ entertainment service	A service that enables the user to play high quality multimedia data transmitted from the outside of home, such as movies, MP3 (MPEG Audio Layer-3), or HDTV (High-Definition TV) on the audio/video devices in home - which is connected via wire or wireless home network, and also enables the user to retrieve the contents inside home from outside, or play games.	HDTV broadcasting, wire/wireless streaming, VoD (Video service on Demand), games, etc.
Home communication/ information service	A service that supports data exchange, the Internet browsing or remote information processing by connecting computers to other computers and printers.	e-mail, Internet search, home shopping, e-government, remote medical treatment, silver care, file sharing, etc.

come stagnation in mature home appliance market by taking advantage of the innovative intelligent appliance technology. They announced independent solutions subsequently to cope with the imminent home network services. Telecom and broadcasting service providers sought to develop new type of services in preparation for converging telecom and media services. Especially broadcasting service providers had necessity to renovate their broadcasting network in preparation of the digital broadcasting. Telecom service providers, on the other hand, were considering new converging business areas such as IPTV (Internet Protocol based TV) services. Contents and solutions providers also intended to capture their share in the new market to take advantage of their competence that has been acquired in the Internet services. Table 2 summarizes industrial participant's service focus for the converging services of home network.

GOVERNMENT POLICY AND DEVELOPMENT PLAN

Korean government had targeted the home network industry as a new growth engine that was driving forces to boost other industries for Korean economics. The home network industry was featured as one of the eight crucial service sectors for next Korean economy, and included as one of the nine IT growth engines constituting so called "IT-839 Strategy" pursued by the Ministry of Information and Communication. The IT-839 Strategy was public policy to expedite the growth of the IT industry through government support (Ministry of Information and Communication, 2005).

In or around 2004, many leading research institutes in Korea and foreign countries forecasted that the home network market in Korea would be created in earnest in 2005 and to be diffused nationwide in 2007. According to their reports, the world home network market was expected to grow at an annual average rate of 19% - 40.7 billion US dollars in 2002, 102.6 billion dollars in 2007, and 162 billion dollars in 2010 (National Information Society Agency, 2003; Ministry of Information and Communication, 2005). Korean home network industry was also expected to achieve a high growth rate (annual average growth rate of 32%) – 2.51 billion dollars in 2002, 11.79 billion dollars in 2007, and 23.45 billion dollars in 2010. Korean government set itself the ambi-

Table 2. Home network service focus of industrial participant

Industrial participant	Home network service focus
Fixed line telecom service provider	• Utilizes the network infrastructure. • Encourages multimedia-oriented contents.
Wireless telecom service provider	• Emphasizes wireless networking and mobility. • Controls using the mobile handset, and emphasizes the automation service.
Broadcasting service provider	• Provides communication/broadcasting convergence service using the CATV (Cable Television) network. • Pursues digital integration technology and economy of scale through the DMC (Digital Media Center).
Home appliance manufacturer	• Utilizes strength of terminal and platform using PLC (Power Line Communication) and media center. • Actively seeks partnerships with construction companies for built-in home appliances.
Construction company	• Value adding for luxurious condominiums and large apartment complexes.
Contents/Solution provider	• Provides solutions, such as set-top boxes, middleware, and digital contents.

tious goal of providing home network services for 10 million households (approximately 61% of the total household) by 2010. The target was to occupy world market share of 15.3% by producing home network related products worth 15.7 billion dollars in 2007 and 28.6 billion dollars in 2010. Experts forecasted home network to be one of the major next-generation export industry. The export volume was estimated as 3.91 billion dollars in 2007 and 7.44 billion dollars in 2010. However, it was known that growth forecast for 2007 was excessively optimistic as of July 2006. The government's original plan to diffuse up to 60% in 2007 was postponed first to 2008 and next to 2010.

Due to the convergent nature of home network industry, many government organizations such as the Ministry of Commerce, Industry and Energy, the Ministry of Information and Communication, and the Ministry of Construction & Transportation are inter-related. The policies to promote the development of the home network industry by related government ministries were as follows.

The Ministry of Construction & Transportation planned to invest 79.3 million dollars in 2004 and 359.8 million dollars in 2008 to encourage the construction with pre-installed LAN and an access network to the Internet, which is so called 'intelligent building'. The ministry also initiated commercialization research center to promote the development of the next-generation home networking technologies. In addition, they planned to create an exemplary home network apartment complex at Masan city in Gyeongsangnamdo province by investing 48 million dollars. Furthermore, the strategy to boost rapid development of the national smart home industry was proposed by reforming the laws and regulations related with construction of the smart homes and buildings (Korea Home Network Industry Association, 2005; Ministry of Construction and Transportation, 2006).

The Ministry of Information and Communication carried out the first and secondary pilot projects to explore standard home network model optimized for the residential environment. The first pilot project in 2004 introduced various converging services adopting front-end technologies, so that the foundations for the home network industry could be suggested. Two consortiums, led by two leading telecom service providers, KT and SK Telecom with the participation of 83 other firms and public agencies, played major roles for the project. They discovered and applied 61 types of exemplary home network services with 38.9 million dollars investment. The prototype of services, consisted of interactive digital television, home automation, health care, home security, information, and entertainment, was offered for 1,300 households in five areas (Seoul Metropolitan area, and other four cities of Busan, Daegu, Gwangju, and Daejeon). In addition, the Ministry of Information and Communication promoted standardization to secure interoperability among communication, broadcasting, and home appliances. They also planned to invest 57.9 million dollars on building smart apartments to provide infrastructure necessary for home network service evolution. However, this goal had not yet been achieved as of July 2006.

The second pilot project, which had been planned to be implemented from 2005 to 2007, focused on development of advanced infrastructure and technologies such as BcN (Broadband convergence Network), IPv6, and ubiquitous networking. However, KT and SK Telecom, the leaders of the first pilot projects up until 2005, decided to expedite full-scale commercial home network services in 2006 instead of further initiating advanced pilot projects. Consequently, the Ministry decided to terminate the existing type of pilot projects in 2006. Afterwards, the Ministry shifted its policy focus to apply exemplary leading public service models, such as remote medical treatment and intelligent home robots.

The home network industry usually involves home automation services such as remote control, home appliance control, and home security, which require the installation of a network and equipment in the early stage of building construction. As such, the role of the Ministry of Construction and Transportation is important, since they can practically affect the strategy of construction companies. The Ministry intended to encourage necessary building infrastructure suitable for providing home network services. To this end, the Ministry revised the Construction Law to add the terms of "intelligent home networks", which mean the LAN based Internet enablement facilities, into the construction facility glossary in May 2006. The Ministry also extended the list of common residential facilities by adding information and cultural facility for the purpose of home network related service provisioning. In addition, the Ministry selected some newly built apartment complexes as the innovative home network service buildings, and encouraged ubiquitous-ready town development in conjunction with Korea National Housing Corporation that is a public agency for house building.

POOR PROGRESS IN KOREAN HOME NETWORK INDUSTRY

After the 1st home network pilot service projects, pursued by the Ministry of Information and Communication, were completed, the two consortiums pointed out the following problems that were revealed from the pilot service provisioning experiences (Byeon, 2006; Ministry of Information and Communication, 2006).

- Equipment and service compatibility were secured within the single pilot project consortium boundary, but inter-working and compatibility between the two consortiums were not properly organized.
- The participants maintained passive attitude towards providing open protocols and networks necessary for home network platforms, home information appliances, and mobile communication. It was due to a sharp conflict of interests among service providers.
- Public services including electronic government, disaster prevention, and crime prevention were provided, but the services lacked necessary inter-working and collaboration with the public and government agencies.
- Inter-working and collaboration with the agencies responsible for social infrastructure, such as gas, electricity, and waterworks were insufficient.

These issues indicated that the distinctive business interests of the service providers, the differences in their profit model, and the lack of openness among the participants prohibited necessary cooperation to enable cross-industrial convergent services and comprehensive value creation. The home network industry encompasses IT base industries such as home information appliances, handsets, chipsets, wire/wireless networks, middleware, and gateway, as well as non-IT industries such as construction, security services, medical services, movies, and education.

We draw upon Wirtz's framework (2001) to analyze the reasons why the progress of the Korean home network industry had been lagging behind the government expectation. The plausible factors that hampered the successful proliferation of the home network services are organized into four categories; customer, technology, business strategy, and policy. The factors organized in Table 3 are interrelated among the factors of three categories; customer, technology, and business strategy. Considering cause and effect relations as well as mutual dependency of the factors across the factors of three categories, we elicited three typical scenarios that could explain the reasons why Korean home network evolution has not yet progressed. We did not consider factors of policy category in developing typical scenarios since they are external factors for business competition.

Table 3. Factors hampering the development of the Korean home network industry

Category	Issues	Factor Number
Customer	Customers didn't regard the service as indispensable.	(factor 1)
	Unsatisfactory value versus cost	(factor 2)
	Customer's low recognition of products	(factor 3)
Technology	Simple product options such as high-speed-Internet-ready apartment only	(factor 4)
	Complicated installation and insufficient indoor wiring or wireless connection	(factor 5)
	Lack of terminal distribution to support full connection of home	(factor 6)
	Technology incompatibility between companies and products, and incomplete standardization	(factor 7)
	Information security issue	(factor 8)
Business strategy	Limited target customer base (intelligent apartment residence, or young housewives)	(factor 9)
	Unclearly defined customer value expectations	(factor 10)
	Lack of business model overcoming the expensive price in early stage	(factor 11)
	Unsatisfactory quality, contents, and reliability due to an absence of cooperation among service providers	(factor 12)
Policy	▪ Multiple ministries and offices involvements in the home network industry ▪ Insufficient deregulation and technical specification enactment ▪ Digital right management issue ▪ Inadequate policy for communication/broadcasting convergence	

- **Scenario 1:** Customers who expected to adopt the home network service in their old house found it difficult to adopt the services because installation was too complicated and no guidelines were provided for indoor wiring or wireless connection (factor 5), and the terminals supporting home network were not easy to get (factor 6). On the contrary, from the service provider's standpoint, it was too risky to bear the expensive cost to be incurred for renovating old digitally-isolated houses in the initial stage of home network deployment (factor 11). In addition, technology incompatibility issues arose among the equipments supplied by the multiple service providers due to the lack of standardization. Moreover, the customers were not expected to get paid off the additional cost which had been incurred for home network installation when they sold the house afterwards. This made the providers cope with customer requirements passively. Further, the service providers could only offer simple product options for new intelligent apartment (factor 4). Hence customer dissatisfied with the value of the service compared to the cost (factor 2).
- **Scenario 2:** Customers residing in a new digital-ready apartment or condominium also were not satisfied with the typical home automation services provided by the built-in intelligent home appliances with connecting terminals. Most customers did not regard the home network enabled automation service as indispensable (factor 1) and felt the cost is too expensive compared to the perceived service value (factor 2). The reason for this was that the service providers narrowly

targeted the service recipients to housewives living in the intelligent apartment (factor 9), and that service product configuration had become too simple (factor 4). While housewives in new apartment complexes might not be technology renovators, the technology renovator group expected advanced service. However, a limited customer target hindered the formation of demand in the early stage of the market. Indeed, companies providing newly built apartments also tried to minimize the cost of expensive equipment such as the indoor facilities needed to use the home network service, refrigerator, and TV. As such, unfortunately, the customer value expectations and value delivery were not matched for those technology innovator group (factor 10), which resulted in the customer's low recognition and poor product awareness (factor 3). That leads to an even greater negative impact on the proliferation of the home network service (factor 6).

- **Scenario 3:** Customers who were considering moving to another apartment had to expect that they would encounter home network service offered by a different service provider. In this case, customers were uncertain whether they could use their existing terminals and home appliances in their new apartment because of technological incompatibility among service providers and products caused by the delayed standardization (factor 7). Consumers would have difficulty in using the same service when they are to move to a new house equipped with different wiring or wireless specifications. Considering Korean's relatively high percentage of moving, an indoor infrastructure and terminal that supported high mobility should be provided through standardization. Although they complained to the service providers for inconvenience, service providers excused that they could not resolve the problem. Indeed, they admitted the reality of insufficient quality, contents, and reliability due to a lack of cooperation among home appliance manufacturers, construction companies, telecom service providers, and contents and solutions providers (factor 12). In addition, when customers moved to a new apartment, they were concerned about the information security also, which influenced them to hesitate about prolonging the service subscription (factor 8).

PARTCIPANTS INTEREST AND CONFLICTS AMONG THEM

In this section, we outlined conflicts among participants with reference to press release data and interviews conducted with industrial participants. Telecom service providers viewed the home network services as the expansion of their existing telecom service business exploiting service platforms which had been provided for the residential service in the household. On the other hand, the home network service was regarded as the new business opportunity to home appliance manufacturers. That is, home appliance manufacturers regarded intelligent home appliance terminals as a driver to induce significant changes in customer behaviors and living patterns, which could lead to new source of revenue through new type appliances selling. Indeed, it could be a new challenge to home appliance service providers in the area of broadcasting and communication field. They could expand their new digital appliance business to include the telecom service areas - something hitherto unattainable.

Telecom service providers expected expansion of terminal market, including home information appliances, and are nurturing potential TV portal market. They tried to diversify capturing in terminal sales, and electronic services such as digital contents distribution and TV commerce. Telecom service providers presumed that the home network equipments were just tools for the home network services, and asserted that appliance manufacturers only needed to provide the interface equipments enabling home network services. As telecom providers had already provided complementary interface equipments, such as telephone modems, DSL (Digital Subscriber Line) modems and, wire & wireless AP (Access Point), for residential telecommunication services, they insisted a platform package for the home network services should be offered from themselves, although home appliance manufacturers and solutions providers strongly opposed to it.

Home appliance manufacturers attempted to take the initiative with home network gateway and indoor connection technologies. They emphasized that the provision of the gateway is most critical for home network services, and that they intended to secure the brand power to provide superior technology based gateway. As such, the home appliance manufacturers competed with each other to capture technical standards in early stage. For example, Samsung Electronics and LG Electronics were in confrontation over the home network middleware standard. In addition to the middleware standard, ascertaining the interoperability among diverse home network middleware technologies is also crucial to proliferation of home network services. However, the technical structure capable of accommodating diverse middlewares has not yet been standardized yet, which also affected the competition among them.

Both the telecom service providers and CATV based broadcasting service providers concentrated their efforts on network advancement. Their strategic focus was on the possession of the highly advanced and evolutionary networks which could be prerequisite for the various home network services including telecom and broadcasting convergence service. As they convicted that home network should be the foundation of evolutionary all IP-based networks, competition on standardization and business conflicts aroused. Consequently, they expected risk taking would be necessary to cope with enormous investment cost, increasing complexity of home appliances, infrastructure development, proper pricing, securing of contents, and cannibalization of their existing markets.

Considering their existing business capabilities, telecom service providers emphasized the convergence services based on communication network, whereas broadcasting service providers emphasized the contents-oriented convergence services. Thus, competition between communications service providers and CATV-based broadcasting service providers over killer broadcasting contents was intense. In this vein, telecom service providers targeted entertainment broadcasting contents as one of killer applications, and tried to cooperate with the terrestrial broadcasting service providers, film studios, and music publishers. They tried to exploit the home network infrastructure to provide the broadcasting contents using interactive TV or media PC, with the addition of interactive communication and interaction. Terrestrial broadcasting contents were heavily emphasized as the core contents of the home network broadcasting & entertainment services, for the home network services in order to evolve from a control/automation services in the initial stage to a broadcasting & entertainment service afterwards. However, telecom service providers and contents/solutions providers conflicted with each other over the absence of a killer application and the necessity for virtuous cycle investment. Indeed, that caused telecom service providers to hesitate to invest on home network services. They perceived that no killer application has appeared in the home network services, and postponed the investment until the application services that would surely expand the market finally emerge.

Telecom service providers considered investment on home network area should prove sufficient ROIs (Return on Investments) since they had made enormous investments on other IT areas such as WiMAX (Worldwide Interoperability for Microwave Access), HSDPA (High Speed Downlink Packet Access), and NGN (Next Generation

Network). Telecom service providers also considered that they have the strongest core capability in the home network industry comparing other industrial participants. Hence they used to invest only after observing the maturity of the market. This passive attitude was also due to the fact that the business area of telecom service providers was mainly restricted to the domestic market. Conversely, the position of the solutions/contents providers was quite different. They planned to enter the overseas market after laying foundations in the domestic home network market.

The competition between home appliance manufacturers was also significant in the area of TV portal. For instance, Samsung Electronics focused on collaboration with telecom service providers, whereas LG Electronics targeted to have associations with Internet portals such as Daum Communications. One of the telecom service oriented consortium attempted to secure profits by inducing large-sized contents providers as their partners, and LG Electronics opted for the strategy of sharing UCC (User-Created Contents). The strategies of the two companies looked different, but it is highly probable that their business models would converge into a similar format, as the former targeted securing of high quality professional contents, whereas the latter targeted the P2P (Peer-to-Peer) market where consumers themselves upload their contents. KT and CATV service providers considered VOD service as one of killer services. They purchased lots of movies, music videos and soap dramas providing the VOD subscribers with access fee charge.

Home appliance manufacturers and chip/PC manufacturers were also intensely competing in the platform and terminal market. Home appliance manufacturers strived to develop, as an enabler of intelligent home appliances terminal and TV portal, various technologies regarding such as operating system, program search, and user interface. However, chip manufacturers and platform vendors had already entered into competition with home appliance manufacturers by taking advantage of their core capability captured through easily upgradeable and highly compatible chip development. Based on this, PC manufacturers, home appliance manufacturers, and chip manufacturers expected the dominant positioning over the appliance manufacturers in the home network service platform areas. The outcome of this struggle would be dependent upon how easily the various contents could be transferred to the consumer via diverse terminals.

Various strategies for the indoor home network technology were also in conflict with each other. For instance, extant fixed-line technologies competed with wireless technology as it underwent rapid evolution. Even competitions were prevalent among wireless technologies with the emergence of wireless local area or personal area communications. Moreover, since each company supported a certain technology by forming a consortium according to its interests, particular technology selection or the integration of these transmission technologies was complicated. Middlewares to secure interoperability among the various technologies were also complicated issue, because many companies competed with heterogeneous technologies and supported the standardization that was beneficial to them.

POLICY IMPLICATIONS FOR HOME NETWORK INDUSTRY GROWTH

Government Direction of Competition Inducement

In the past, Korean government announced several competition inducing policies to boost the information and communication industry growth. Since 1990's, the government promoted some selected telecom providers to induce competition with the existing dominant service provider in the areas such as mobile communication and long distance and international telephone services (Ministry of Information and Communication, 2006).

Unlikely to past experiences, recent digital convergence services are distinctive since they are fusions of technologies, industries, and policies. Due to industrial convergent nature of the home network, services such as communication, broadcasting, wired and wireless services are provided at the same time by many industrial participants. Under this industrial merging environment, providers are not easily distinguished and competition spreads across the industry borders. Therefore, the government policy to boost the industry through inducing competition, based on the single industry dichotomy, would not be valid. Instead of inducing competition based on inflexible separation of particular industry, it would be more desirable to direct the competition policy into encouraging strategic alliances among the industrial participants engaged in home network industry.

Government Initiative on Technology And Infrastructure

It is effective that the government leads technology development initiative at the early stage of innovative service inception in the areas of information and telecom service industry which is characterized as rapid technology evolution and heavy investment on necessary infrastructure. The government could initiate various public investments to provide the technology infrastructure to induce flexible entry of the business firms into the new business areas. The government policy to invest on specific core technology will also encourage firms to enter the market more aggressively. The policy has been effective for the success of CDMA (Code Division Multiple Access) and the high-speed Internet propagation in Korea. However, unlikely to the case of CDMA, Korean government encountered difficulties in initiating technology development for the home network industry. Indeed, the scope of technology was too wide due to the convergent nature of the home network services - wherein different types of information were prevalent and processed by various digital equipments. For instance, there were many kinds of digital signals transmitted in the house, which included communication, broadcasting, and home appliance controls.

Considering all theses, the government initiated pilot project through formation of consortiums for the demonstration of the innovative exemplary applications of home network service. The policy intended that prototype of interactive application services could be effective to induce the construction of the home network infrastructure in public schools, public agency offices, hospitals, and so on. For example, a public agency could offer electronic civil services at home, which in turn should affect the consumer recognition of the convenience of the home network services. Initial experience of the convenient pilot services should lead to increased demand for the overall home network services. The lessons indicated that the government technology leading will necessarily focus on the compatibility and convergence aspect of technology, which enables compatible interface among different technologies. It is the premise of strategic alliance that this study espouses on.

Government Market Regulation

Government market regulation means that government intervenes in the market and limits the actions of enterprises or individuals to ensure justice in economic society. The government regulation is effective to resolve the issue of economic power concentration, imbalance among income groups and regions, and to enhance basic rights. In the information and telecom industry, the most important market regulation of Korean government was the classification of information and telecom business areas in order to reduce the total social cost through avoiding duplicate investments and to boost market competition among the same service providing group.

Considering the convergent nature of the services, the regulation policy needs to be redirected to reflect a new service classification instead of specific single industry based classifications as the past. The new classification could classify the group of industries as transmission service providers (telecom and broadcasting industry) and program service providers which supply re-organized and customized programs to end users. As such, it could be the first step to identify industry groups to reflect convergent characteristics of home network services. Secondly, the regulation should be devised with the premise to cultivate domestic industry. That is, the regulation should not become an obstacle to the growth of the convergence industry.

Standardization and Certification

Standardization involves establishment of common technical specifications, and procedures to ensure the standards be properly followed by the industrial participants for the customers to exploit the services conveniently. Thus, if the standardization properly works, it should contribute to both the service provider's profit generation and the consumer's welfare. However, service providers, who try to occupy dominant market share by capturing their own standard, could restrict the benefits of customers. Customers expect to get the service at a moderate price through competition among the suppliers and popularization of the services. Thus, conflict will arise between the service providers attempting early occupation of the market with their preferred proprietary standard, and customers who could be better off in case the market allows flexible inter-connection with any type of technologies.

It was pointed out that the lack of government efforts for setting up unified certification procedure necessary to establish technical standards of home network services, resulted in the slow progress of standardization and home network service propagation. Certification of technical standard is the process of deciding whether the core technology, which has been developed by the company over a long period, can be distributed to the commercial market or not. The standardization process verifies the validity and reliability of the developed technology. Although a couple of private organizations were organized by standardization agencies and private companies, the issues of establishing technical standards and compatibility satisfying various service providers have not yet been resolved.

As a first step of expediting standardization in home network industry, the certification process needs to be established. A national certification center establishment could be imperative for this purpose. The center would manage the standards of the home network technologies, and the certification procedures. Such a certification center could award the certification for the home network equipment and specific service, thereby play the role as a catalyst to achieve compatibility among the products. The certificate would include certified mark that ensures QoS (Quality of Service) for digital signal sharing among home appliances, indoor security devices, and wire and wireless telecommunication. Academic groups in research institutes and universities could work together to prepare necessary certification procedures required for the home network service enablement.

CONCLUSION

In this chapter, we reviewed a brief history of the Korean home network service evolution from the inception of the industry to recent converging service developments. Despite the nation-wide efforts for proliferating growth of the home network services as an industry to lead the Korean economy growth, the current progress does not fulfill the expectation. Basically, our analysis indicated that current home network service level is not mature to satisfy customer's demanding expectations. Three home network service scenarios elicited from customer perspective reveal that conflicts of business interests among the industrial participants comprising convergent home network services are the root causes of poor progress of home network services. To resolve these conflicts, four policy implications are proposed from digital convergence perspective. The suggestions include government roles on new direction of competition inducement, initiative for technology and infrastructure, market regulation, and standardization. In the subsequent chapter, we discuss the managerial implications for the participating parties to elaborate the profitable business model in the home network service area.

REFERENCES

Byeon, S.-B (2006). *Customer Needs Analysis and Service Provisioning Method for the Home Network Service*. Korea Radio Station Management Agency.

Kim, C.-H (2005). *Trend of Digital Home Network Development*. Korea Electronics Technology Institute.

Korea Home Network Industries Association (2005). *Status of Local and Overseas Home Network Industry and Future Development Strategy*. Jinhan M & B.

Ministry of Construction & Transportation (2006). from http://www.moct.go.kr.

Ministry of Information and Communication (2005). *Plan for Intelligent Home Network Service Distribution and Industry Nurturing Plan*, Government Report.

Ministry of Information and Communication (2006). from http://www.mic.go.kr.

National Information Society Agency (2003). *Information Technology Advancement, Planning Report on Next-generation IT Growth Engine (Home Network Part)*, IITA.

Wirtz, B. W. (2001). Reconfiguration of Value Chains in Converging Media and Communication Market. *Long Range Planning*, *34*, 489-506.

ENDNOTE

[1] This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-041-D00885).

Chapter XV
Digital Convergence and Home Network Services in Korea:
Part 2 – Business Models and Strategic Alliances[1]

Hyun-Soo Han
Hanyang University, South Korea

Heesang Lee
Sungkyunkwan University, South Korea

Yeong -Wha Sawng
Electronic and Telecommunications Research Institute, South Korea

ABSTRACT

In this chapter, we proposed managerial implications and strategic alliance opportunities of the home network service providers in Korea. Based on in-depth analysis of recent industrial status of home network services in Korea, we first analyzed business models of each interested parties participating converging services of home network industry. Next, in order to complement the weakness of participant's business models, strategic business alliance opportunities and managerial implications were drawn upon resource-based view (RBV) theories. Subsequently, strategic propositions for the success of the home network are suggested at the firm level, with respect to three main converging home network services. Finally, we offer future directions and considerations to expedite home network service evolution. The findings and propositions suggested in this study could be exploited for further research on other digital convergence services.

INTRODUCTION

Home network is a digital convergence service on the base of integrated computer information technologies, telecom services, and home appliances in residential areas. Despite the promising expectations and significant impact of the home network on Korean economy as one of the major growth driver, the home network industry is lagging behind the initial government and experts forecasts.

In Part 1 of this study, we investigated why the progress of the home network industry fell short of expectations from the perspectives of customer value expectations and business interests of industrial participants. The analysis revealed the business conflicts among the participants were unavoidable due to the convergent nature of the home network. Subsequently, some government policy implications were proposed to enhance the customer service level through effectively resolving the conflicts among the participants.

In this chapter, we continue the discussion at firm level to provide the strategic implications for home network industry. Based on the premise that effective collaboration to resolve the conflicts among the participants should be necessary for the digital convergent services, our analysis focuses on exploring business alliance opportunities among the participants in the home network industry. We investigate business model of each participant to analyze the value proposition and strategic target. Subsequently, adopting resource capability theories (Barney, 1991), strategic alliance opportunities are explored. Finally, other considerations for the home network service expansion are discussed in light of digital convergence to offer strategic propositions at the firm level.

BUSINESS MODEL ANALYSIS OF HOME NETWORK

This section provides comparative analysis of the business models of the various participants in the home network industry with reference to press release data and interviews conducted with industrial participants. In addition, literatures reporting the past experience of the Korean home network (Byeon, 2006; Kim, 2005; Korea Home Network Industries Association, 2005; Ministry of Construction & Transportation, 2006; Ministry of Information and Communication, 2005; Ministry of Information and Communication, 2006) were also referenced for our analysis.

Business model means a concept and implementation system that secures and uses an enterprise's resources to provide better customer value than that of the competitors. The theory of business model draws more attention as the proliferation of Internet based companies and verification necessity of the technological and commercial feasibility. Industrial participants of home network services were diverse due to inherent nature of convergence from many industrial sectors. In our study, the home network services were classified into three types; home control & automation service, home broadcasting & entertainment service, and home communication & information service. The companies participate in home network services were from various industrial sectors such as telecom, broadcasting, home appliances, construction, and contents and solution providers. The details for service classification and participant analysis were presented in Part 1 of this study. For the purpose of analyzing business models of home network services, we summarize the components of the business model as in Table 1 by referring to the study of Afuah and Tucci (2006).

Table 1. Business model components

Component	Description
Customer value	A value provided to the customer
Customer segment	Target customer of the service, to whom the value is provided
Scope of services and products	Scope of the products and services provided to the market
Price	Pricing method and strategy
Source of revenue	Pattern of the profit that is achieved while providing the product
Business activities	Activities that should be carried out to achieve the value
Implementation	Organizational structure, human resources, and system needed to carry out business activities
Capability	Capability of providing the value in a better way
Sustainability and growth opportunity	Degree to which a competitive advantage over other companies can be maintained

Business Model of the Telecom Service Provider

Telecom service providers are capable to provide the most competitive home network service. They could utilize the existing subscriber base, marketing capability, and distribution channels that they have cultivated while running the Internet access and/or mobile communication business. Telecom service wished to concentrate on digital contents and related additional services like VoD (Video service on Demand), MoD (Mobile service on Demand), Internet-based or mobile broadcasting services, remote medical care and education. In particular, telecom service providers tried to utilize existing wire/wireless communication network infrastructure. In addition, telecom service providers emphasized the experience they have acquired in providing seven days twenty-four hours communication and access service request interactively.

Paradoxically, these advantages induced them to take a passive attitude regarding investment in the home network business, compared with other service providers. Fixed line telecom service providers believed that their core asset is pre-installed access network infrastructure including FTTH (Fiber to the Home) which enables high-speed Internet access up to the home. Wireless telecom service providers regarded their strength as the capability to provide wireless communication across the indoor and outdoor network environments. They had past business experiences in wireless voice and data communication and pre-distributed mobile handsets to a large number of people. Since telecom service providers had strong revenue generation from existing telecommunication services, they intended to expand their business with only moderately pricing level of home network services that required minimal additional investment on the infrastructure. They were hesitant to invest massively on new home network service area considering the uncertainty of revenue generation of home network. Fixed line telecom service providers wished to concentrate on enhanced multimedia contents services with high QoS (Quality of Service) exploiting their core capability of providing broadband high-speed communication. On the other hand wireless telecom service providers emphasized indoor mobility and portability to provide digital contents services and home network automation control services utilizing their omni-present handsets widely distributed to the existing subscribers. In sum, business model components that should be suitable for the telecom service provider are summarized as Table 2.

Table 2. Business model of the telecom service provider

Component	Description
Customer value	• Provide innovative and ease of use home network services exploiting connected and automated home environment
Customer segment	• Customers subscribing high-speed Internet access service for fixed line telecom service providers • Customers subscribing mobile telecom services for wireless telecom service providers
Scope of services and products	• All home network services • Multimedia contents emphasis for fixed line telecom service provider • Mobility and portability of the service/contents for wireless telecom service providers
Price	• Access fee + one time usage fee
Source of revenue	• Savings on indoor network construction cost and frequent usage inducement
Business activities	• Service portfolio upgrade and expanding compatibility/scalability of various digital contents and applications.
Implementation	• Ensure reliable and secure service provisioning • Cost-effective operation, maintenance, management, and systems
Capability	• Experience of seven days twenty-four hours uninterrupted telecom service
Sustainability and growth opportunity	• Economically establishing post home network services against initial terminal embedding investment by construction companies or home appliance manufacturers

Business Model of the Broadcasting Service Provider

Broadcasting service providers had the richest range of experience as a contents integrator. They also had good customer's expectations for ease of use, which could be symbolized by TV and remote controller experiences, fitted well into the characteristics of the home network services. As such, broadcasting service providers expected their strategic focus towards securing the convenience and simplicity of the user interface in the home network services. Rather than public broadcasting service providers, CATV (Cable TV) broadcasting service providers got enough infrastructures for home network services exploiting their interactive communication & broadcasting convergent service.

Broadcasting service providers emphasized home broadcasting & entertainment services that share multimedia contents, rather than home control & automation services. CATV service providers had already distributed the set-top boxes that could be used as the gateways of home network services. The strategy emphasizing home broadcasting & entertainment services is based on the capability to provide communication & broadcasting convergence services by upgrading the cable modem and the set-top box of the established CATV network.

CATV broadcasting service providers have competed with telecom service providers in high-speed Internet access service area. The competition were intense since telecom service providers would also provide IPTV (Internet Protocol based TV) service via high-speed Internet, and broadcasting service providers would offer triple bundling service that aggregate broadcasting, telephone, and high-speed Internet access via the CATV network. The CATV broadcasting service providers have intended to utilize the household Internet access infrastructure of the CATV network to maximal extent.

At present, broadcasting service providers in Korea are in comparatively weak position compared to telecom service providers for the following reasons. First, the revenues of major telecom service providers (KT and SK Telecom) are much higher than that CATV System Operators (SO). Second, broadcasting service providers

Table 3. Business model of the broadcasting service provider

Component	Description
Customer value	• Provides various contents and home network services, focused on easy of use, as an extension to existing convergent services including TV and Internet
Customer segment	• All CATV service subscribers and potential customers
Scope of services and products	• All home network services
Price	• Access fee + one time usage fee
Source of revenue	• Convenience and savings on the cost of controlling indoor service equipment and terminals
Business activities	• Endeavor to diverse control automation, communication, and information service product development • Expand compatibility and scalability of various home network applications
Implementation	• Interactivity service provisioning, cost effective service development, operation • Management of organization, human resources, and system
Capability	• Secure competitiveness from economy of scale in terms of service provision and customer interaction cost • Marketing capability to cope with intense competition
Sustainability and growth possibility	• Maintains convenient usability while expanding communication, information, control, and automation service

have not yet been exposed to intense competition due to their regional monopoly status. It could be a weakness considering the dynamics of the home network industry. Finally, because CATV has not yet been fully digitalized, they are not familiar with providing the interactive network services. However, recent strategic movement of cooperation and integration among CATV service providers centered on DMC (Digital Media Center) has rendered itself to obtain economy of scale necessary to digital broadcasting. Considering all these, the business model components that are suitable for the broadcasting service providers are described in Table 3.

Business Model of the Construction Company

Construction companies has accumulated rich experience in providing the Internet access enabled apartments and condominiums in 1990s. They also had experiences in construction of indoor network and platform environment construction, which were cultivated while constructing intelligent office buildings. Indeed, the real competitive strength of construction companies was their capability to create a fundamental infrastructure necessary for home network service provision, which other participants did not possess. Moreover, construction companies had abundant knowledge of behavioral patterns of customers at home. It was an important asset for designing home network services compatible with home resident's daily behavioral patterns. The another potential competitive strength of construction companies was in their capability cultivated for the preparation of the u-City (ubiquitous City) project which enables ubiquitous applications at a city level not just limited to a building or apartment complex level. Based on these competences, they emphasized built-in type of service provisioning for the home network services.

It is imperative to ascertain home network equipment compatibility. For this purpose, construction companies sought to invite telecom service providers or broadcasting service providers as the home access network provider, and home appliance manufacturers as the home network platform and terminal provider. In addition, construction companies wanted partnership with Systems Integrators (SI), Network Integrators (NI), and contents & solution providers. The partnership with SI, NI, and contents/solutions providers could complement their lack of experience in providing seven days twenty-four hours digital contents and telecom services.

Construction companies were relatively free from cost limitation, as they could provide all-in-one home network services as a value added built-in services for luxurious condominiums or large apartment complexes. Actually, the cost portion of home network provisioning was relatively small in total construction cost. In most cases, customers do not seriously compare the cost and the value of the home network service. However, there was a limitation in that relatively luxurious houses equipped with home network services was estimated as less than 10% of the gross national hosing market in Korea. Further, they are not capable to maintain the services, for instance, upgrading the middleware for home networks. Aggregating all these, the business model components that are suitable for construction companies are summarized in Table 4.

Business Model of the Home Appliance Manufacturer

Big Korean home appliance manufacturers sought an opportunity for additional revenue generation from effective integration of home network services with existing home electronic appliances. Within the big picture of an intelligent home in the ubiquitous society, home network services connect home appliances with digital contents, broadcasting, and communication. Thus, propagation of home network services should provide product innovation or differentiation opportunities for the appliance manufacturers. Adding connectivity to digital appliances would render them for more revenue generating strategic opportunities. Moreover, the expertise of analyzing consumer needs as well as innovative product design capability, cultivated through fierce global competition in the past, should be the core competence of home network service provisioning.

These strategic maneuvers drove them to concentrate on capturing industry standard in indoor networking and platforms including solution and

Table 4. Business model of the construction company

Component	Description
Customer value	• Builds up a digital home that reflects the residential environment and compatible with residential behavior practice of the customer
Customer segment	• Mainly for the customers inhabiting the apartment and condominium
Scope of services and products	• All home network services
Pricing	• Infrastructure construction cost + access fee
Source of revenue	• Saves on the cost of indoor network establishment and service alliance, and implements a innovative services compatible with behavioral pattern in the house
Business activities	• Independent implementation of the control and automation service • Provide various communication, information, broadcasting, and entertainment services with business partners
Implementation	• House construction, and operation, maintenance, organization, human resources, and information system to secure service scalability and cost-effectiveness
Capability	• Secures the integration and combination capability of various services including digital contents
Sustainability and growth opportunity	• Capability of securing compatibility and coping with the aftermarket of more proliferated home network services

Table 5. Business model of the home appliance manufacturer

Component	Description
Customer value	• Provides a home network service using connected, intelligent, and convenient home information appliances
Customer segment	• All customers who use electronic appliance at home
Scope of services and products	• All home network services
Pricing	• Equipment purchase cost + access fee
Source of revenue	• Secures dominance in the terminal market through platform standardization, and saves the cost of indoor networking environment construction
Business activities	• Develops a rich control/automation service product, and expands the compatibility and scalability of various applications
Implementation	• Effective partnership with construction companies, solution vendors, and digital contents providers
Capability	• Capability of identifying customer needs, functional innovation of electronic appliances based on this, and securing service stability
Sustainability and growth opportunity	• Relative advantages in providing the home network services, compared with the telecom service provider and broadcasting service provider

contents services for home network. On the other hand, they were actively looking for partnerships with construction companies in order to include built-in electronic home appliances into the basic configuration of new house infrastructure of the construction companies. To ensure uninterrupted stability of the services, home appliance manufacturers also pursued partnerships with SIs, NIs, and hardware & software providers. In sum, the business model components that are most suited to home appliance manufacturers are summarized in Table 5.

Business Model of the Contents and Solution Provider

Contents providers supplied the digital contents for the broadcasting and entertainment service via fixed and mobile telecommunication service providers. Evolution of the home network services could be a new opportunity to expand their line of business. Solution providers that had competence in control and automation service tried to expand their business from developing new software solutions for home network services in the areas of communication, digital broadcasting and entertainment services. However, since most Korean contents and solution providers were small and medium sized, and specialized in some particular solution, technology, and contents areas. Hence, they had close partnership with large broadcasting companies and telecom giants. To cope with the new home network service evolution and further to exploit the trend as their business expansion opportunities, enhanced cooperation with the relevant parties participating home network industry would be effective. While captivating core competence, they need to pursue diverse ways of cooperating with telecom service providers, broadcasting service providers, construction companies, and home appliance manufacturers. In this perspective, the business model components that could be suitable for contents and solutions providers are summarized in Table 6.

Table 6. Business model of the contents/solution provider

Component	Description
Customer value	• Enrich home network services to make it more convenient and fun by providing better contents delivery and ease to use environment
Customer segment	• All customers who use the home network service
Scope of services and products	• Contents and solution for home network services
Pricing	• Purchase cost + access fee + one-time usage fee
Source of revenue	• Discovery of specialized applications at a moderate development cost
Business activities	• Contribute to extend the scope of home network services and customer base by developing various control, automation, and customizable service products and digital contents
Implementation	• Planning, development, organization, human resources, and information systems to develop customized service and to provide rich contents
Capability	• Secure specialized development and provisioning capability to promptly cope with diverse customer demands and customized service
Sustainability and growth opportunity	• Establish active partnership to overcome disadvantages relatively small company size and captive market share

STRATEGIC ALLIANCE OPPORTUNITIES IN HOME NETWORK INDUSTRY

The value chain of the broadcasting industry, wire/wireless telecom industry, and Internet access industry has been disaggregated and a new type of converging value chain is now being formed. As the value chain is re-configured for the home network industry, the roles assigned to each service provider and the profit sharing structure of the service providers need to be reorganized. While each distinctive service provider could generate the value as per their own business model in their independent business domain, the roles and value addition of the service providers should be reshaped for delivering value in the home network environment.

In general, a firm expands its business domain through strategic alliance with the industrial partners. However, in the home network industry, the roles could be overlapping among the participants to arise conflict, which did not positively affect the customer value. Unlikely to traditional competition among the suppliers providing the similar services, it would not be beneficial to customers that the each participant comprising the segment of convergent value chain of the home network competes with each other. The competition among the participants of the home network industry caused by incongruent business models asserting their own interests could result in staggering the market growth. As such, we draw upon resource based view theory to derive propositions necessary to encourage strategic alliance.

Resource-Based View (RBV) Perspective

Corporate resources include all the assets, capabilities, abilities, organization processes, and knowledge that can be controlled by the enterprise. Resource-based view (RBV) theory typically explains firm's competitive advantage and collaborations adopting the concepts of resource heterogeneity and non-mobility (Peteraf, 1993; Wernerfelt, 1984). Resource heterogeneity understands the firm as a collection of production resources, and this collection of production resources is discriminated with each other. Non-mobility resources possessed by the firm are characterized as to be difficult to get imitated by other competitors and inelastic to supply. Heterogeneous and non-mobility resources possessed by the firm can be the potential source of capturing competitive advantage. That is, those resources can enable the firm to sustain its competitiveness or to alleviate threats from the competitors.

We organized the necessary resource capabilities, adopting the framework of Mills & Platts (2003), required to provide the home network services in Table 7. Indeed, the convergent nature of home network services requires aggregation of diverse resource capabilities cultivated by each industrial participant comprising the value chain. For instance, home appliance manufacturers have solid fundamental resources such as factories, facilities, and raw materials. The telecom service provider has the capability to provide twenty-four hours seven days non-stop service. Also, the construction companies are familiar with collaborative relationship with subcontractors for construction business. Collaborative aggregation of each distinctive resource capabilities from each participant should be necessary to provide the competitive home network services.

Resource Complementarity and Strategic Alliance

RBV perspective emphasizes creation of new capability through dynamic combination of distinctive resources. The advantage gained from resources combination will be stronger in case each distinctive resource capability can be complementary as a whole (Gullati, 1998). In early stage of the market growth, the firm's competitive advantage is usually captured through its unique

Table 7. Resource capabilities in home network industry

Resource typology	Distinctive resource for home network service
Tangible resource	• Physical resources, products, patents and others that are related with the home network service
Knowledge resources (knowledge and experience)	• Control & automation service: capability for production or sourcing of semiconductors and controller • Broadcasting & entertainment service: knowledge and experiences of providing broadcasting, games, and educational contents • Communication & information service: knowledge and experiences of developing and providing digital information and telecom service • Implicit know-how on design and development of innovative product and services
Procedural and systems resource	• Control & automation service: techniques and know-how on procurement and sales of control & automation products and platforms • Broadcasting & entertainment service: systems and process know-how in distributing broadcasting and advertising • Communication & information service: Systems and process know-how on sales, operation, and billing the communication & information service
Organization culture and value	• Customer-oriented culture, profit maximization and sales mindset on home network services
Network resource	• Control & automation service: Partner network for providing control, automation parts and solutions • Broadcasting & entertainment service: Contents provider network and public brand recognition • Communication & information service: Information & communication industry supplier group partner networking and telecom service brand
Dynamic capability	• Control & automation service: sustaining development of innovative product to keep up with advanced technology • Broadcasting & entertainment service: continuous update and renovation of digital contents • Communication & information service: keep on adapting to advanced technologies of digital information services

and valuable technology capability. However, the initial competitiveness of the firm will be less stable as the market matures to establish industrial standard of the technology. Especially, when the technology could be imitable easily, the technology-based competitive advantage captured in early stage will be lessened sooner. To sustain competitiveness, Harrison et al. (1991) and Tripsas (1997) proposed the importance of complementarity leverage effect newly secured from effective combination of distinctive assets cultivated from other firms. The impact of resource complementarity is stronger in case the secured assets cannot be imitated with easy and can be cultivated only through long period of time. For instance, Amazon.com complemented their distribution capability through obtaining offline distribution centers which, otherwise, required long time to establish. The four types of resource complementarity are organized as follows (Amit & Zott, 2001).

- Product and service complementarity: Creates a new value from bundling the products or services, e.g. one-stop shopping, customer support service.
- Offline asset complementarity: Generates synergy through complementary utilization of offline assets, e.g. brand reputation, distribution channel.
- Business activity complementarity: Generates synergy through complementary utilization of the firm's business activities such as supply chain integration.
- Technology complementarity: Creates a new value through complementary combination of technologies, e.g. image processing and Internet technology combined video conference system.

The fundamental convergent nature of home network services suggests significant value enhancement through effective exploitation of resource complementarities. Business model of the home network services could be more solid to provide better customer value through dynamic combination of core competences of industry participants. The strategic advantage from business alliance can be ascertained by the synergy captured through resource complementarity as described in Table 8. Based on resource complementarity, we propose the cooperation and alliance

Table 8. Value creation opportunities from resource complementarity

Complementarity type	Value creation opportunities
Product and Service	• Complementarity through integrating home communication and broadcasting service • Complementarity gained from controlling indoor devices using the mobile handset • Complementarity effect through extending after service maintenance efforts of home appliance manufacturer to home automation control
Offline asset	• Synergy from capturing brand reputation of the home appliance manufacturer and telecom service provider • Synergy obtained from complementary utilization of distribution channel of broadcasting, digital contents, and communications
Business activity	• Synergy effect by utilizing the maintenance organization of the telecom service provider and the broadcasting service provider • Complementarity effect by integrating the supply chain activity with digital contents creation from contents provider
Technology	• Technology integration of information processing and communication • Interactive broadcasting and t-commerce through complementary exploitation of communication and broadcasting technology • New value creation such as remote medical treatment enabled by effective integration of communication and control technologies

strategy for home control & automation services, home broadcasting & entertainment services, and home communication & information services.

Control & Automation Services

The technological breakthrough of control & automation services was made in the areas of factory automation and electronic control. Though home appliance manufacturers and solution providers have the core capability in this area, they need to complement other distinctive technologies which they lacked in to provide the home network services. Due to the indoor characteristics of the home network services, the home control & automation devices require complicated wire and wireless connection inside the house. Home appliance manufacturers and solution providers need to collaborate with telecom service providers or broadcasting service providers, since the control & automation of home network equipments are expected to be operated remotely via wired or wireless Internet, .

Core capabilities of home appliance manufacturers and solutions providers are knowledge and system assets. Knowledge assets include expertise in production and utilization of hardware and software for control and automation, and system assets include the know-how on platform implementation, operation, sales, and after service. If the capabilities of providing wired or wireless Internet technology from telecom or broadcasting service providers are added, the value of control and automation services could be enhanced since the services could be operable from remote. This is the case of synergy from product and service complementarity. In addition, while the control & automation services provided by home appliance manufacturers or solutions providers is basically onsite and intermittent, the service provisioning of the telecom and broadcasting service providers is continuous, user-included, and remote. As such, the combined capabilities can enable customer support of home control & automation services to be made instantly as per the customer's request from remote.

Broadcasting & Entertainment Services

Although contents quality is most critical for home broadcasting and entertainment services, seamless connectivity of digital contents supply chain is necessary to properly distribute high quality contents to customers. That is, the value of broadcasting & entertainment services of home network can be ascertained when the core capability of broadcasting & entertainment contents providers is effectively combined with the competence of communication, broadcasting, and information processing technologies. Moreover, it is not a simple one way delivery of broadcasting & entertainment contents, but interactive and real-time services. It implies that the specialized broadcasting & entertainment contents provider needs to establish a cooperative partnership with broadcasting service providers, telecom service providers, and home appliance manufacturers at the same time. Moreover, the effective alliance could provide the advertisement opportunity to secure economic viability of the services. When these capabilities are effectively combined, the broadcasting & entertainment services in the home network would be representative technology complementary type of services.

It would be plausible for the broadcasting service provider to lead the broadcasting & entertainment services because they are biggest source of contents creation. Also, System Operators (SOs) and Multiple System Operators (MSOs) could take the role to exploit the recent digitalization trend of CATV and DMC facilities. In addition, small and medium sized contents providers could actively participate in this area by engaging in a strategic alliance with broadcasting service providers, telecom service providers, and home appliance manufacturers. These players, however, need to closely collaborate with other service providers such as telecom service providers and home appliance manufacturers. However, both the telecom service providers and broadcasting service providers need to prevent the cannibalization of their existing service markets. For example, if one service provider offers low priced bundling services of communication, contents, entertainment, and automation to their existing customers, it could affect the existing revenue structure of the voice, data, and multimedia communication services.

Communication & Information Services

Communication services in the home network extend the existing communication between peoples up to the communication among terminals. Extended communication among terminals is characterized as the automatic provisioning of information collected from various connected terminals in the house. Telecom service providers are in a better position to provide this extended communication services to take the leading role. However, they will provide the better value in case they form a strategic alliance and cooperate with other industrial participants. In particular, they need alliance with home appliance manufacturers and solution providers to complement the capabilities of processing, controlling, and monitoring the status information generated by the terminal equipments.

The communication & information services in the home network industry include active provision of Internet-based communication & information service to households, and ubiquitous applications furthermore. For example, a ubiquitous application such as monitoring refrigerator inventory to alarm shortage of food items can be possible in case the home network infrastructure is fully utilized. Forming an alliance is required to develop this kind of integrated and innovative services. The other potential ubiquitous applications could be the remote medical and education services. In providing theses services, effective collaborations with relevant parties are necessary to secure their professionally vested rights.

DISCUSSIONS

In this study, we proposed managerial implications of the business models of each industrial participants and strategic alliance opportunities among them. The results indicated that development of home network services should be pursued from industry convergent perspective to exploit resource complementarity for competitive service provision to customers.

The other considerations for the home network service expansion in Korea include the overcoming of digital divide. The home network services provided to intelligent apartments or buildings cannot be the same as those for rural individual house. Firm's distinctive strategy, as well as government policy, is needed to be addressed for the rural houses and low income households. Finally, since that many ubiquitous applications could be implemented as the home network services, next-generation ubiquitous services need to get effectively combined with the scope of the home network service. As such, in order to provide the full-fledged home network services, the next-generation ubiquitous applications need to be discovered sooner along with the cross industry strategic alliances.

REFERENCES

Afuah, A., & Tucci, C. L. (2006). *Internet Business Models and Strategies: Text and Cases.* McGraw Hill.

Amit, R., & Zott, C. (2001). Value Creation in E-Business. *Strategic Management Journal, 22*(6/7), 493-520.

Barney, J. B. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99-120.

Byeon, S.-B (2006). *Customer Needs Analysis and Service Provisioning Method for the Home Network Service*. Korea Radio Station Management Agency.

Gullati, R. (1998). Alliances and Networks. *Strategic Management Journal*, *19*(4), 293-317.

Harrison, J. S., Hitt, M. A., Hoskisson, R. E., & Ireland, R. D. (1991). Resource Complementarity in Business Combination: Extending the Logic to Organizational Alliances, *Journal of Management*, *27*(6), 173-190.

Kim, C.-H (2005). *Trend of Digital Home Network Development*, Korea Electronics Technology Institute.

Korea Home Network Industries Association (2005). In M. & B. Jinhan (Eds.), *Status of Local and Overseas Home Network Industry and Future Development Strategy*.

Mills, J., & Platts, K. (2003). Applying Resource Based Theory, Methods, Outcomes and Utility for Managers, *International Journal of Operations and Production Management*, *23*(2), 148-166.

Ministry of Construction & Transportation (2006). from http://www.moct.go.kr.

Ministry of Information and Communication (2005). Plan for Intelligent Home Network Service Distribution and Industry Nurturing Plan.

Ministry of Information and Communication (2006). from http://www.mic.go.kr.

Peteraf, M. A. (1993). The cornerstone of competitive advantage: a resource based view. *Strategic Management Journal*, *3*(3), 179-191.

Tripsa, M. (1997). Unraveling the Process of Creative Destruction: Complementary Assets and Incumbent Survival in the Typesetter Industry. *Strategic Management Journal*, *18* (1), 119-142.

Wernerfelt, B. (1984). A Resource-based View of the Firm. *Strategic Management Journal*, *5*(2), 171-180.

KEY TERMS

Business Model: Description of means and methods a firm employs to earn the revenue projected in its plans.

CATV SO (CATV System Operator): A cable company that serves only one system or community.

Complementarity: Material or good whose use is interrelated with the use of an associated or paired good such that a demand for one generates demand for the other.

FTTH (Fiber to the Home): The installation of optical fiber from a telephone switch directly into the subscriber's home.

MoD (Mobile service on Demand): An umbrella term for a wide set of technologies and companies whose common goal is to enable individuals to select multimedia contents from a central server for viewing on a television or computer screen.

MSO (Multi System Operator): A cable company that serves multiple cable television systems or community.

u-City (Ubiquitous City): High-tech future city built on ubiquitous infrastructures, technologies and services to improve life quality and raise city value by systemizing administrative functions and processes of the city.

Section III
Telecommunications Business Management

Chapter XVI
Mixing and Matching Organizational Network Legitimacy Practices to China's Telecommunication Market

Brian Low
University of Western Sydney, Australia

Wesley J. Johnston
Georgia State University, USA

ABSTRACT

Despite massive investment, few foreign firms doing business in China's telecommunication sector can claim success. Such failures can be explained by their inability to acquire organizational legitimacy. Legitimacy is garnered when firms successfully market their competencies to key stakeholders, by conforming to regulative processes, institutional norms and cognitive meanings within the environment. But which competencies should a firm push to develop its legitimacy and to whom should they be targeted? While firms can push its in a variety of different ways, there is no guarantee of success in a network of embedded, interdependent stakeholders. Drawing on institutional theory, this chapter presents a case research based framework to describe an iterative and incremental process to help firms manage their network legitimacy. Using this framework, the chapter examines the complementary assets and legitimacy orientations of these firms with network stakeholders, and the resulting inter-partner initiatives and alliances.

INTRODUCTION

Every month, senior strategist of foreign firms operating in China's telecommunication sector, analyze and develop scenarios to better understand and manage it. Armed with insight and knowledge from research and intelligence, consulting assignments and industry briefings, they plot the path to survivability and profitability. On the surface, there is a calm of normalcy and the noise of "business as usual." But as the complexity for developing business strategies increases, there is a sense of urgency and anxiety. This is not an unusual scene because in a telecommunication sector undergoing structural reforms, these enterprises are forced to operate in a range of settings, often subject to different "rules of the game" (North, 1990).

Despite having gone through an arduous process of soul searching in relation to the development of the telecommunication sector, policies remains underpinned by concern over maintaining control of a strategic interest and the protection of key stakeholders vested interests within the state–controlled economy. While the Government is committed to structural reforms, policies continue to be affected by powerful stakeholders with vested interests. Rather than stepping back and letting the market operate, "government and industry protectionism" have instead resulted in a raft of conflicting and ambiguous policies announcements.

To deal with the sector's uncertainty and complexity, strategists need to take a more expansive view of the environment. The competitive reality in a transitional market economy is that success depends upon whole constellations of stakeholders and foreign firm's ability to garner organizational legitimacy by meeting the norms and expectations of these stakeholders. The determination of legitimacy is made by the stakeholders to which the organization must be responsive, and on whom it is dependent for survival (Kumar and Das, 2007). Legitimacy is thus concerned with how firm and / or stakeholders perceived one another, and how they may co-operate in future inter-partner alliances and programs. Legitimacy in turn facilitates access to other firm's resources and activities needed in the production and transformation of industrial goods and services.

Our concern in this chapter is with network legitimacy. In industrial networks, firms are linked together by their performance of industrial activities (e.g. marketing, exporting, production, logistic), employing or consuming various types of resources (e.g. R&D, financial, brand equity, knowledge) to produce other resources. Over time and many interactions, firms develop tangible relationships that are connected together to form a "quasi-organization" (Hakansson and Ford, 2000). The system of these interconnected relations makes up the business network, in which the firm is embedded (Wilkinson and Young, 2000). Firms operate in the context of these business relations and networks, affecting the nature and outcomes of their actions. Hence, resources owned and activities performed by the firm are not as important as their relative attractiveness as perceived by others. This attractiveness, at a point in time, and over time, stipulates the firm's network legitimacy.

This chapter contributes to the existing knowledge on organizational legitimacy in three important ways. First, we extend our understanding of organizational legitimacy by introducing the concept of network legitimacy, i.e. legitimacy in industrial networks. Second, we examine the process of network legitimacy by proposing a theoretical model, based on the interaction among: (1) the firm's reputation and network characteristics; (2) legitimating initiatives, and (3) legitimacy outcome. This model is depicted in Figure 1. The reputation of the firm is crucial in any legitimacy seeking exercise, especially among foreign firms without any previous, local business dealings. This exercise will also be influenced by the characteristics of the network in which they are seeking legitimacy.

Figure 1. Organizational legitimacy – a network perspective

Network characteristics include aspect of business relationships, network positions, and network structure. Both the reputation of the foreign firm and network characteristics influence the types of legitimating initiatives vis-à-vis their complementarity of resource and activity mix and targeted legitimacy orientations, required to initiate, and subsequently manage their network legitimacy. Third, we conduct an inductive interpretative case analysis of Motorola China, allowing us to determine the measure of support of our concepts, the proposed interactions compromising the model, and addressing managerial concerns.

BACKGROUND

Legitimation refers to the social justification of an actor or activity, such that the actor or activity is publicly validated or endorsed (Perrow, 1961). The process of social validation involves recognition of a distinctive competency possessed or role-played by the organization in providing a good or service to relevant stakeholders (Dacin et al, 2007). Institutional theorists have examined this process via the relationships between these organizations and their environmental context (DiMaggio and Powell, 1983; Scott 1995). Their basic tenets have been that when studying the adoption of particular organizational strategies, organizations need to achieve and maintain environmental legitimacy in order to survive. Firms survive by conforming to institutional norms within the environment, in turn earning them legitimacy (DiMaggio and Powell, 1983; Hillman and Wan, 2005). What makes managing legitimacy particularly difficult is that it exists on a continuum and a unanimous agreement of all parties is not necessary for legitimacy to exist (Kumar and Das, 2007).

Within a broader international contextual setting, network studies have examined how firms in different stages of their internationalization have managed their relationships with business and political actors (Child & Tse, 2001; Havila, Johanson, & Thilenius, 2004). As institutional settings [including any consequential contextual change] within which networks evolve, these studies have found that political institutions and states often provided "a framework of rules and regulations within which private actors [business enterprises] have to play" (Salmi, 1995, p. 68). We note particularly the existence of these frameworks in China's transition from a centrally controlled to a market economy (Low and Johnston, 2005; Nolan 2001). This transition has resulted in changes to the nation's business ideologies, with profound impacts on the reconfigurations of business networks across many politically salient and sensitive sectors, including telecommunications.

Foreign firms operating in China's telecommunication sector therefore face the daunting challenge of identifying these changing ideologies, including changes in changing expectations as existing networks are reconfigured. It is an onerous and complex task, given the nation's deeply embedded institutional norms, values and governances. The lack of transparency processes in the formulation of policies, conflicting national imperatives, absence of a uniform copyright and intellectual law, makes the task seemingly impossible. Ambiguous policies are often announced, with frequent mid-stream changes. A case in point is the endless delays over the past 5 years in rolling out China's home-grown TD-SCDMA 3G mobile phone technology. These imperfections are perhaps not unexpected as the telecommunication sector transforms itself from what has been a centrally controlled sector to a semi – capitalist juggernaut. As Chen Jinqiao (2003), the director of China's MII's Institute of Telecom Policy put it:

…The current set-up, though not perfect, reflects three "unique" conditions in China: government-run corporations will remain the main driver for the national economy, serious disparity in local markets, and telecom reform must proceed in step with the economic system which is also controlled by the government.

In choosing their strategies, foreign firms in China no longer share an environment to which they have become accustomed to in the West. Many have failed not because of poor products or a lack of resources but rather because of inappropriate or ineffective efforts to build legitimacy (Ahlstrorn and Bruton, 2001). Others simply find themselves in a hopeless situation in China (Vanhonacker, 1997), once they are labeled illegitimate. Firms with higher levels legitimacy management experience will, however, have a more precise view on what is needed to pursue and navigate a successful legitimacy seeking agenda. Prior legitimacy experience in turn results in more opportunities to enter into future inter-partner alliances, presumably due to the development of the firm's reputation. In short, foreign firms that manage to survive long enough in China are more likely to have conformed to legitimacy pressures while those that do not, will not survive.

The key question for the strategists is: what are the most important combinations and complementarities of resource and activity mix required to push their network legitimacy? Should these be limited to economic driven inter-partner initiatives like R&D, minority equity position, or engagement in manufacturing, sourcing, and re-exporting programs? Or should they also include public welfare initiatives including educational programs and regional developments? What would their basis of their claims to legitimacy i.e. legitimacy orientations? To whom should they be targeted, in order to stage and manage their legitimating initiatives? We attempt to address these questions, starting with a brief review of the concept of network legitimacy.

ANALYZING NETWORK CHARACTERISTICS AND LEGITIMACY

Foreign firms seeking to initiate their network legitimacy agenda generally need to undertake an analysis of the network characteristic, much like market analysis. The initial task is to identify the focal firm with whom a business relationship is planned. This firm, however, is likely to have ties to other firms, through past and current relational exchanges. The relationships that exist between them would have evolved over time and over many interactions. As a result, relationships between these firms create interdependency yet limit their ability to act independently, including planned new relationship with foreign firms seeking entry into the network.

Interdependency between these firms also creates some form of an inter-organizational network structure, with varying degree of structuredness. The more structured the network, the more difficult it is for foreign firms to initiate legitimacy seeking initiatives. This is because in any given structure, relationships between firms meant different functions for each of these firms, given their roles. These roles are defined by these firms' relationships such that the cumulative effects of these relationships are reflected in the structural network positions, in terms of the degree of interdependency. Hence, network structures are as much a process as they are structure, being constantly shaped and reshaped by the actions of firms who are in turn constrained by the structural network position in which they find themselves (Nohria, 1992, Sydow and Windeler, 1998; Dittrich et al 2006).

These distinctions are important because one of the network characteristics is the recognition that firms within the network need to project a unified view vis-à-vis of the larger environment in order to garner external legitimacy. Unity is needed to ensure compliance with the expectations of powerful external stakeholders that may include regulatory bodies, institutional authorities, non-government organizations and the central government. They exert a powerful influence in charting the sector's policies, awarding tenders, setting technology standards and benchmarks, approving collaborative alliances, and determining the scope of initial public offerings (IPO), among others. There are, therefore, strong incentives for these firms to comply.

Compliance is achieved through commonly used strategies and practices that often emerge from the interactions of firms and other stakeholders within the network (Edelman, 1992). This process of 'collective making of meaning' within the network, shaped by the 'politics of propriety, trust, and awareness' (Nielsen and Rao, 1987, p. 523) determine the survivability and profitability of these firms and the network. While disagreements between firms are not unexpected, a properly functioning internal network dynamics that revolves around cooperation and trust is crucial in order to maintain inter-firm network legitimacy. Firms cooperate with each other because they each can provide the capabilities and resources they lack to be competitive.

If inter-firm network legitimacy has to change, it is because the network no longer possesses the resources needed to meet external stakeholders' changing expectations. Major external structural, regulatory, and institutional changes may force these stakeholders' hands. The network and the firms within it may have to source for new resources, to ensure their survival and profitability. The powerful forces of conservatism that have so far successfully internalized these external expectations will come under increasing pressure. Questions will be raised on the roles of these firms and the network, in the transformation and production of new industrial goods and services. This scenario often provides the pretext and the context in which relationships with foreign firms will be sought.

Expanding along this line of reasoning, inter-firm relationships stems primarily from efforts to match their resource and activity complementarities i.e. firm's assets, to achieve economic benefits. Foreign firms seeking entry into the network seek the same benefits. However, because not all networks have the same benefits, there is priority among foreign firms in seeking out local firms with complimentary assets. The greater the assets complementarities, the greater are the benefits from combining their assets under the rubric of inter-partner alliances and partnerships. Information about potential local, network partners is therefore a key resource. So is the ability to assess and predict the complementarities of their assets.

ORGANIZATIONAL REPUTATION AND AFFECTS ON LEGITIMACY

Beside the tangible elements that foreign firms bring to the relationship which include the resources they own and the activities they perform, they also bring intangible elements such as knowledge, culture and reputation. For these firms without any prior dealings with local firms, its reputation is both relevant and crucial, especially at the early stage of their legitimacy seeking agenda. There are no precedents or sales histories to study to determine the compatibilities of these firms. And with increased competition among foreign firms seeking a presence in the local market, assessment of their compatibilities is likely to be based on the reputation of these foreign firms.

A favourable organizational reputation thus acts as a core intangible resource that creates competitive advantage when competitors are not able to match the prestige and esteem it creates (Shrum and Wuthnow, 1988). Reputation represents the affective or emotional reaction to a firm and is defined as the overall emotive estimation of a firm by its constituents (Fombrun, 1996: 37). Central to reputation is therefore a comparison of organizations to determine their relative standing. For any two organizations, they will either have the same reputation or, more likely, one will have a better reputation than the other (Deephouse and Carter, 2005, pp. 331). Significantly, lack of reputation might mean rejection of the foreign firms' legitimacy seeking agenda from the outset.

Once a preliminary assessment of a foreign firm's reputation is confirmed by the local partners and / or stakeholders, the question is whether over time, the initial assessment is warranted? This spawns many related questions: Does the complimentary of tangible and intangible elements satisfy the needs of all network stakeholders or only a select few? Is the culture of the foreign firm compatible with those of the network? What impact, if any, would it have on current internal network dynamics?

Along these lines of questions, we like to think that the firm's reputation at the initial stage of their legitimacy seeking agenda afford the stakeholders an opportunity to evaluate the firm as part of the social construction process. Over time, reputation may or may not be attained, depending on the role of the foreign firm and their contributions in helping the network to attain and sustain superior outcomes. Issues concerning the various legitimating aspects of maintaining reputation as part of the social construction process, therefore, cannot be ignored. The interface between reputation and legitimacy is a work-in-progress. This is something that foreign firms should be thinking about, even if they cannot predict the outcome with changing stakeholders' expectations.

RESEARCH METHODOLOGY

Various characteristics associated with China's telecommunication sector lend it to readily observable trends and developments. The central government involvement, the nation's ambitions of becoming a global technological power, structural reforms, the nation's obligatory WTO commitments and the protection of stakeholders' interests, are some of the characteristics that presents the sector as an ideal "critical" or "polar" case where the "process of interest is readily observable" (Eisenhart, 1989, p. 537). These observations and analysis are gauged from publicly available documents through internet sites, company reports, trade journals, and commentary by industry researchers and analysts. Through an inductive, interpretative, content analysis of these materials and through methodological contextualization via industry experience, we were able to provide some empirical descriptions and generalizations. We achieved this through the use of case research, allowing us to study interactions between our proposed variables and the empirical evidence, albeit exploratory. We next examine this process through an inductive, interpretative case analysis of the sector's key legitimacy orientations, followed by an analysis of Motorola (China) Electronics Ltd.

MANAGING LEGITIMACY ORIENTATIONS AND MATCHING COMPLIMENTARY ASSETS

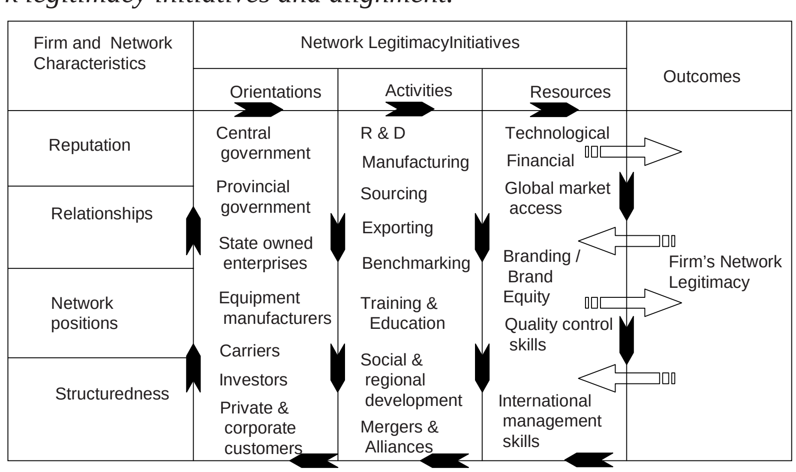
The quest for network legitimacy is motivated by the fundamental need of foreign firms to garner stakeholders support. Success depends on their ability to match their assets i.e. resource and activity mix, with those of the key stakeholders. The formulation of this mix is, however, influenced by the type of network ties that exist among these stakeholders, each with their interpretation of what constitute network legitimacy. Legitimacy may be pragmatic, moral and cognitive in nature (Suchman, 1995). It may also be managerial and technical in nature (Ruef and Scott; 1998).

In a business-to-business marketing setting in China, legitimacy may also be state-oriented, partner-oriented, customer-oriented and global-oriented legitimacy nature (Wei and Bello, 2004, Low et al, 2007). In this manuscript, we apply these orientations and foreign firms' matching complimentary assets to the context of firm's pursuit of network legitimacy in China's telecommunication market. Anecdotal evidence suggests that network legitimacy has been built through a mixture of global, partner, customer and provincial legitimacy orientations on the back of the firm's reputation and an understanding of network characteristics. These orientations are manifested through technological leadership, investment in R&D activities, local manufacturing, sourcing and exporting, education and social development programs and initiatives. These orientations do not come together at one. They evolve in the context of the overall legitimacy development process, through trial and learning. They require constant monitoring and realignments, in response to network disruptions where the actions and reactions of the key stakeholders in the network must be considered to ensure a "dynamic fit" with network disruptions, as shown in Figure 2.

Global Oriented Legitimacy (GOL)

Foreign firms with a GOL is generally informed about what is going on in China's transition from a developing nation to an industrialized state, on the back of structural and institutional reforms in the telecommunication sector. As a result, they ensure that their competencies match those of the key stakeholders, especially those with the responsibility to realize the nation's aspiration. Because many of these stakeholders are controlled by the government, and because the government has significant power to approve telecommuni-

Figure 2. Network legitimacy initiatives and alignment.



cation projects and allocate resources, a GOL is crucial in helping smooth foreign firms network legitimacy agenda. For instance, in China's attempt to develop TD-SCDMA, its own version of 3G, foreign firms with global 3G technological innovations and influence in setting technology standards, would be in a strong position to cultivate partnerships with key stakeholders.

Provincial and State Oriented Legitimacy (PASOL)

It is generally acknowledged among telecommunication analysts and commentators that China's telecommunication reforms have not generally provided the same level of incentives for businesses operating in different provinces. Businesses and households in different provinces also do not have the same level of access to telecommunication services. Strong institutional gaps also arise from a lack of market, technical and political support through the central government's diminished role. These developments provide the ideal context for foreign businesses to work collaboratively with local authorities, and perhaps, contribute to the provinces economic and social development. A PASOL that is underpin by foreign firms social and moral obligations in underdeveloped regions, while not economically feasible, may position these firms as good corporate citizens: an important legitimacy attribute.

Partner-Orientated Legitimacy (POL)

A POL refers to foreign firm's ability to work with local equipment manufacturers. Historically, these partnerships, particularly manufacturing, mirror the sector's national competitive development policy (Low, 2005). This has resulted in a market structure where local manufacturers like Huawei, Zhongxing and Datang holds substantial market power, often supported by government–led polices. While lauded in the formative years of the sector's evolution, these policies are no longer tenable in an increasingly liberalized telecommunication sector. In order to survive, these manufacturers are now pursuing aggressive international market expansion programs and R&D alliances. Companies with a POL, especially those with competencies in global market access and technology innovations may find it easier to pursue their legitimacy agenda, through exporting and R&D alliances with these manufacturers.

Customer Oriented Legitimacy (COL)

Foreign companies have historically faced high entry barriers in dealing with local carriers like China Telecom and China Mobile. These carriers are also much less transparent and have a more bureaucratic background compared with their international carriers. But as pressures mounts on these carriers to access global funds through initial public offer (IPO), the investment community will expect more transparency and less bureaucracy from these carriers. Any preference these carriers have for sourcing local equipment from local manufacturers would therefore be sorely tested, especially with increasing pressures to increase revenue yields, through reduced operating costs and capital expenditure. Foreign firms, especially those involved in the manufacturing of telecommunication equipments, can capitalize on these developments by adopting a COL through a mix of superior, innovative cost-effective technological products, services and solutions.

In conclusion, an integral part of these firms' network legitimacy initiatives is their ability to match their complementary assets and legitimacy orientations with those of key stakeholders. These orientations and complementarities do not come together at one. They evolve over time, influenced largely by the actions, reactions, and expectations of the network community of stakeholders. Their influence requires constant monitoring. For many foreign firms, the question is how to manage legitimacy orientations and matching complimentary assets to maximize the probability of success in achieving network legitimacy? We next examine this in the context of Motorola China.

NETWORK LEGITIMACY IN CHINA'S TELECOMMUNICATIONS MARKET: THE CASE OF MOTOROLA (CHINA) ELECTRONICS LTD

Motorola (China) Electronics Ltd. (hereafter MC) opened its first office in China in 1987, and in 1992, established a wholly owned manufacturing facility in Tianjin (80 kms east of Beijing). Since 1987, it has invested over US$3.6B in building local manufacturing and R&D capabilities, including over US$450 million having been invested in 18 research and development (R&D) centers. The company has over 3000 personnel focusing on R&D activities alone, the largest presence of any global company in China. MC also has over 30 technology cooperation projects with more than 20 Chinese partners. MC operations now consist of a holding company in Beijing, a factory, eight joint ventures, 18 R&D facilities and 26 sales offices throughout China. The accumulative sales volume of MC reached US$58.1 billion, and its production of cell phone output in China account 34 percent of the company's total worldwide, according to Ruey Bin Kao, president of MC (2006).

MC has performed impressively in what has been until recently a highly protected and regulated telecommunication sector. Their success is due in large part to their ability to articulate their legitimacy orientations and complementarities of assets mix to stakeholders. They recognized and conformed to commonly used legitimacy strategies and practices deemed acceptable to stakeholders. Its success culminated in the nation's 2005 recognition and award of MC with Best Corporate Citizenship Practice and the Most Influential Multinational in China. We next examine in some details some of MC's legitimacy orientations and its complementarities of assets distinctive and how these helped establish the firm's network legitimacy.

Legitimacy, Melding and Fusion of Co-Specialized Resources

As one of the founding member of the TD-SCDMA forum, Motorola has one of the industry's most comprehensive silicon portfolios. The historical role it has played and the resource portfolio it possesses culminated in the signing of an MOU with Datang Mobile in 2004. This MOU is particularly significant because Datang Mobile is the current leading provider of mobile communication equipment in China and the co-inventor and principal driver of TD-SCDMA, China's version of its own 3G technology. Datang Mobile owns this core technology as well as driving the access network.

MC's legitimacy in TD-SCDMA technology predates another earlier relationship it had with China Unicom, the country's second largest and the world's third largest mobile operator, dating back to 2001. MC also has a long-standing relationship with China Mobile, the world's largest wireless operator, culminating in the selection in 2004 of Motorola's Global Telecom Solutions Sector (GTSS) by China Mobile to expand its Global System for Mobile (GSM) communications network.

Not only has MC moved with the right technology and at roughly the right time, it has also managed to do so in conjunction with the two largest mobile operators in China. Together with the relationship MC has with Datang Mobile, MC appears to have carefully crafted and leveraged its global wireless technological legitimacy, by choosing complementary partners through inter-firm sharing knowledge routines that have evolved over time. This has happened despite the disruptive character of emerging wireless technologies that makes extrapolations of their eventual market prospects a futile exercise. As

Adrian Nemcek (2002), executive vice president of Motorola Inc once remarked:

As a global leader of CDMA RF solutions, Motorola is proud to play an integral role in Unicom's CDMA network upgrade. These contracts further strengthen our long-term partnership with China Unicom as well as the Sino-US trade relationship.

Legitimacy, Commitment and R&D Integration

MC has also pitched its technological legitimacy to include significant investment in R&D activities aimed at helping third party developers, such as the creation of the Application Center of Excellence (ACOE) in 2004. According to Patrick Kuong from Motorola North Asia, ACOE provides enhanced cooperation opportunities not only with carriers, but also with strategic solution providers and 3rd party developers in the telecom industry. In 2006, MC announced the opening of its Broadband Wireless China Research Center in Beijing. The center will focus on researching key technologies for future broadband wireless systems and helping develop global standards. Technical expertise will be brought to Motorola business teams in China, driving research programs with customers and universities in China. The center also plans to work with the FuTURE Mobile Communication Forum, an open and international, non-government and non-profitable communication organization. Motorola is a founding member of the FuTURE forum. Similar research commitments such as Motorola Labs that focus on innovative technology for Motorola products and services worldwide generally reflect an increase capacity of local stakeholders to learn from foreign partners.

There is now an increasing expectation that China should not only be seen as a manufacturing and sourcing hub but also a hub in science and technology. As noted in a recent Organization for Economic Co-Operation and Development report, "the time when foreign investors invested in China only to take advantage of cheap manufacturing platforms is over." MC's commitment in local R&D has the potential to make MC the core of a new giant integrated telecommunications business. It places the company at the forefront in upstream and downstream innovation with some of the major stakeholders in the network.

Going into 2008 and beyond, MC had developed into a corporate giant that continues to leverage its technological legitimacy in the way existing industrial networks may be shaped and disrupted, and playing a significant role in the transformation of old industrial production, procurement and consumption in China's telecommunication sector. A fertile source of new legitimacies includes software developers, universities, non-government, non-profitable organizations, technology transfer organizations and research institutes.

Legitimacy and the Nation's Ambitions

In his book, Nolan (2001) has posited that all previously successful industrializing countries have used some form of industrial policy to construct large, indigenous globally competitive firms. The growth of such firms has generally required substantial investment by the central government to develop indigenous large businesses that can challenge the global giants (ibid). MC can assist China realize these ambitions, particularly in wireless technology and handsets. This assistance is what China has been looking for, but seldom acknowledged until recently. The road ahead will not be easy. China is coming from a position of lateness. China's desire to be in the big league of technology leaders is unlikely to materialize for at least a decade. As Jinpei Cheng, China's vice minister of science and technology noted:

China is still in the tier-four rank of scientifically marginal nations, whereas the United States and Japan are tier-one science superpowers. We hope to reach the next tier, to become a 'strong nation' in basic research in the coming 20 years.

Historically, the government has, and will continue to play, a central role in generating high rates of equity investment and stimulating technical progress. They will continue to rely heavily on foreign companies like Motorola to seed the country with R&D facilities. They also see these companies as a catalyst for moving China up a few rungs on the R&D ladder. Significant opportunities existed and continue to exist for MC to help accelerate China's emergence as a technology superpower particularly given MC's proven track record in China. As Crag Watts (2003), an analyst with Norson Telecom Consulting puts it:

There is a lot of trust on the Chinese side because Motorola has put a lot of investment into the country.

And again as vice premier, Zeng PeiYan (2007) recently commented, when expressing his appreciation of the remarkable achievements Motorola has made in the past twenty years since it entered the Chinese market and its contribution to the development of the China's telecommunications sector:

The company should strive for a win-win situation with Chinese partners, and make new contribution to the development of overall China-U.S. economic and trade relations.

Legitimacy and the Prospect of the Local Market

China has one of the world's largest networks. It is also the world largest producer of handsets. The largest mobile phone company in the world is China Mobile, in terms of the number of subscribers. As of July 2006, China has 366 million fixed line subscribers and 431 million mobile customers, making China the largest telecommunications market in the world. China's number of Internet users topped 137 million by the end of 2006, an increase of 23.4% from a year before and 162 million by June 2007, making China the second largest Internet user after the United States, according to China's Ministry of Information Industry (MII). Showing no signs of abating, pressure mounts for local equipment manufacturers and carriers to satisfy demand. This offers new opportunities for foreign telecommunication companies keen to further legitimize their network existence, especially when preferences for sourcing local equipment from indigenous manufacturers diminishes in a deregulated telecommunications sector.

Faced with a booming domestic market, the government has also embarked on colossal manufacturing and sourcing programs. This offers the possibility for large revenue streams, something that has not been lost on MC. Strategizing the market in many different ways, but ultimately with two very important constants – local manufacturing and sourcing, MC has been an active participant in these programs. For instance, while Motorola sell $5.7 billion worth of products in China, it also exports $3.6 billion from China, including $1 billion in international procurement from local suppliers. China, according to Christina Li, a Beijing-based spokeswoman for Motorola, is a winning base of production and sourcing for the company.

As a major benefactor of these programs, MC is also the only foreign company among 19 manufacturers in China to receive a license to manufacture CDMA handsets in the country. As Kao's noted:

Facing these golden opportunities, Motorola will continue to invest in China to develop strategic partnerships with Chinese telecom companies [carriers and manufacturers], bringing to consumers smarter technologies and making their lives simpler.

And as Chinese Vice - Premier Zeng Peiyan eloquently describe:

The hope is that Motorola deepens its participation in China's informatization process, and strengthens research and development to turn out more products suitable for the Chinese market.

Legitimacy and the Public Welfare Imperatives

As previously noted, China aspire to join the big league of world technology leaders. This is not going to happen anytime soon because of their low investment in research and development compared with global telecommunication giants (Low, 2005). While there is high incentive to increase R&D spending, there is an even greater incentive for MC to help train its citizens. Local manufacturers could then become more global and more powerful, developing their technical expertise and their global brand name.

The establishment of Motorola University (MU) in 1988 is Motorola's contribution towards helping China's realize its "big league" aspirations. Indeed, MU was largely created in direct response to China's emerging market driven economy and hence, the need to train its employees. With centers in Shanghai, Guangzhou and Tianjin, MU has also developed many programs targeting senior and midlevel managers from the company's suppliers, strategic partners, state-owned enterprises and customers. MU provides value-added services, the key one of which is delivering educational technology and know-how to China, not only to Motorola's own employees, but also to many of the state-owned universities and enterprises.

Since 1998, and in cooperation with the State Development and Reform Commission of China, Motorola has offered management-training courses to more than 4000 managers from 1000 state-owned companies in 26 provinces across the country. Now more than ever, China needs more Western trained managers to fill the country's WTO entry requirements. According to Lois Webster (2002), former director of MU's China operations from 1996-1998:

Motorola University China has played a very significant business development role for Motorola's operations in China. This confidence in based on Motorola's deep rooted investment in China and its close co-operation and partnership with the Chinese Government and enterprises.

MC has also developed a complete and diversified corporate citizenship program that focuses on educational support and spans over Project Hope – helping school drop – outs to return to schools, higher education, and state – owned enterprise training, environmental protection, and poverty alleviation. In recent years, the company has encouraged its employees' active involvement in volunteer services and has launched a number of public – welfare initiatives in a well – planned and well – organized manner.

In summary, Motorola's network legitimacy has been built on the back of the firm's reputation, and an understanding of the characteristics of the networks, through a mixture of global, partner, customer and provincial legitimacy orientations. These orientations are manifested through the company's global technological leadership, investment in local R&D activities, local manufacturing, sourcing and exporting, education and social development programs and initiatives. These legitimacy orientations do not come together at one. Instead, they require constant monitoring and realignments in response to network disruptions. They evolve over time, where the actions and reactions of the network stakeholders of carriers, manufacturers, software developers, customers, state and provincial governments must be considered in order to ensure a "dynamic fit" with network disruptions, as shown in Figure 2.

In 2006, Southern Weekend, an influential Chinese Business Newspaper, rated Motorola (and by extension MC) the number two spot in its survey of Most Influential Fortune 500 Companies operating in China, and highest in a list of Best-Performing Fortune 500 Companies. It cited financial performance, investment in China, regional contributions, corporate social responsibility, employment of local staff, local sourcing, environmental protection, involvement in public welfare, corporate governance and brand recognition as criteria. We like to think that these accolades is a measure of its success in achieving its network legitimacy by mixing and matching its resources, activities and legitimacy orientations with key stakeholders within the network and external to the network.

FUTURE TRENDS, ISSUES AND CHALLENGES

How to manage a foreign firm's network legitimacy given the conflicting expectations of stakeholders in China's telecommunication sector is at the heart of this manuscript. Our proposed model does not eliminate the uncertainties and complexities inherent in foreign firms' attempts in garnering internal and external network legitimacy: these remain as real as ever. It does however provide some insight on how they could reduce these uncertainties. As important as these insights are, foreign firms confront four major challenges in their attempt to manage and navigate their network legitimacy.

First, in a network of interdependent stakeholders, it is insufficient to target only one stakeholder at a point in time. Often, by not linking its legitimacy to, and the potential impact they may have on other stakeholders, their efforts may be compromised. Assessment of these efforts therefore cannot be made without a rich understanding of the legitimacy of individual stakeholders in the network and the network itself.

Second, there is heterogeneity in identifiable legitimacy initiatives these firms can undertake. These initiatives also occur with limited cognitive ability. As a result, a key objective for foreign firms is to develop linkages with significant network counterparts who act as an information and knowledge guardian and processor (Low and Johnston, 2005). The firm can, with greater certainty, then predict which legitimacy initiatives will work and whether they reflects a good proportion of what the stakeholders want.

Third, legitimizing initiatives that results in long-term inter-partner alliances usually involves a series of short-term improvisation, incremental actions. This is because while foreign firms pursue their legitimacy initiatives by shifting their orientations, resource and activity mix across the network, stakeholders in the network may also be shifting theirs. Both the foreign firms and their local partners also face considerable pressures in drawing upon their internal and external resources and overcoming their internal and external resistance, in implementing these actions.

Fourth, foreign firms need to adopt a process orientation as part of their operational paradigm to grasp the challenges identified previously. The focus is not only with the way these networks looks, but with explaining changes to this network, of which they are a part of. Critically, this involves monitoring shifts in government policies, their impact on inter-firm network dynamics, and the network garnering of external legitimacy. Legitimacy outcomes, in this way, is regarded as nothing more than the temporary and transient effects of foreign firms network legitimacy initiatives that involves constant realignment needed to ensure the relevancy of a firm's legitimacy agenda, and hence the need for a process orientation.

CONCLUSION

In industrial networks, firms are linked together by their performance of industrial activities, employing or consuming various types of resources to produce other resources, thus capturing the essence of firm's network legitimacy. In this chapter, we propose that network legitimacy play a central role in the management of foreign firm's success in China's telecommunications market. However, there is heterogeneity in identifiable legitimacy orientations firms can bring to the vast network of key but interdependent stakeholders. There is also considerable difficult in establishing the firm's complement of resource and activity mix with these stakeholders.

A key feature of our proposal is that these heterogeneities and complementarities manifest in greater pressures upon foreign firms in managing their network legitimacy, especially given the ambiguity of context. Ambiguity arises principally from stakeholders with conflicting interests in a tightly knit network community. Not all of these stakeholders have the same needs. Neither would they necessarily accept that such needs could not be met in a variety of different ways. Foreign firms' success in managing their legitimacy therefore requires effective exploration and balancing of appropriate legitimacy orientations and assets complementarities to demonstrate commitment to the interests of these stakeholders, and acquiring legitimacy from them.

In the telecommunication sector, foreign firms' legitimacy emanates first and foremost from the development and commercialization of innovative technological solutions with key stakeholders within the network, including local equipment manufacturers, carriers, software developers, and investors. They must also consider external stakeholders like regulatory bodies, non-government organizations, regional and central government because of their control over stakeholders within the network, many of whom are state-operated enterprises. External stakeholders, in particular the central government, also play a crucial role in formulating the sector's policies and their impact and contributions to the nation's economic ambitions, awarding tenders, determining technology standards, and approving collaborative alliances. Together with stakeholders within the network, foreign firms must project a unified view vis-à-vis of the larger environment to garner external legitimacy, by meeting the expectations of these external stakeholders. And in a nation that is increasingly concerned with social inequality, foreign firms' social and public welfare legitimating initiatives are just as important as their technology and related economic-driven legitimating initiatives, in driving their legitimacy seeking agenda.

Companies like MC appear to have grasped the many legitimacy attributes and nuances through its investment in research and development and manufacturing, particularly in wireless technology. MC has also launched a series of social, educational, regional development programs and public welfare initiatives, taking the company beyond the realm of economically oriented only legitimating initiatives and relationships. By taking what they do well already in technological innovations and manufacturing and combining it with its societal responsibilities, the company has capitalized on its well-established network legitimacy. A balance, then, of both building its future legitimacy and exploiting its past legitimacy is seen as essential. There is no necessity to be locked into any one particular set of legitimacy orientations and complementarities of resource and activity mix. The focus instead, is, constantly aligning these orientations and complementarities with the demands of stakeholders within the network and external stakeholders, in particular the central government. By conforming to these demands, Motorola have garnered their network legitimacy.

In conclusion, we therefore find the proposed interactions among concepts comprising the model depicted in Figure 1 to have a measure of some empirical support, albeit a speculative based inductive interpretation of archival and contemporary data. Future conceptual frameworks that seek to deal with network legitimacy would however do well to take into account these interactions. In a transitional Chinese telecommunication sector, the setting should prove ideal.

REFERENCES

Ahlstrorn, D., & Bruton, G. (2001). Learning from successful local private firms in China: Establishing legitimacy. *Academy of Management Executive*, *15*(4), 72.

Chen, J. (2003). *Profiting from benchmarks: Reasons behind China's rush to develop standards*. From ChinaTechNews.com. April.

Child, J., & Tse, D. (2001). China's transition and its implications for international business. *Journal of International Business Studies*, *32*, 1-23.

Dacin, M. T., Oliver, C., & Roy, J. J. (2007). The legitimacy of strategic alliances: An institutional perspective. *Strategic Management Journal*, *28*, 169-187.

Deephouse, D. L., & Carter, S. M. (2005). An examination of differences between organizational legitimacy and organizational reputation. *Journal of Management Studies*, *42*(2), pp. 329-360.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147 – 160.

Dittrich, K., Jsapers, F., Valk, W., & Wynstra, F. (2006). Dealing with dualities. *Industrial Marketing Management*, *35*, 792-796.

Edelman, L. B. (1992). Legal ambiguity and symbolic structures: organizational mediation of civil rights law. *American Journal of Sociology*, *97*(6), 1531–76.

Eisenhart, K. M. (1989). Building theories from case research. *Academy of Management Review*, *14*(4), 532-550.

Fombrun, C. J. (1996). *Reputation: Realizing value from the corporate image*. Boston, MA: Harvard Business School Press.

Hakansson, H., & Ford, D. (2000). How should companies interact in business network? *Journal of Business Research*, *55*(2), 133-139.

Havila, V., Johansson, J., & Thilenius, P. (2004). International business – relationship triads. *International Marketing Review*, *21*(2), 172-186.

Hillman, A. J., & Wan, W. P. (2005). The determinants' of MNE subsidiaries' political strategies: Evidence of institutional duality. *Journal of International Business Studies*, *36*, 322-340.

Jinpei, C. (2002). *China research and development goes into orbit*. http://www.my-esm.com/article

Kumar, R., & Das, T. K. (2007). Interpartner legitimacy in the Alliance Development Process. *Journal of Management Studies*, 1-29.

Low, B. (2005). The evolution of China's telecommunication equipment market: A contextual, analytical framework. *Journal of Business and Industrial Marketing*, *20*(2), 99-108.

Low, B., & Johnston, W. (2005). Managing ambiguous policies in China's telecommunication market. *Journal of Asia Pacific Business*, *6*(4), 5-30.

Low, B., Johnston, W., & Wang, J. (2007). Securing network legitimacy in China's telecommunication market. *Journal of Business and Industrial Marketing*, *22*(2), 97-106.

Nemcek, A. (2002). *Motorola receives US$446 millions CDMA 1X contracts from China Unicom*. Motorola Press Release.

Nielsen, E. H., & Rao, M. V. H. (1987). The strategy-legitimacy nexus: A thick description'. *Academy of Management Review*, *12*, 523–33.

Nohria, N. (1992). *Is a Network Perspective a Useful Way of Studying Organizations?* In Nohria and Eccles (Eds.), *Networks and Organizations*. Boston, MA: Harvard Business School Press.

Nolan, P. (2001). *China and the global economy*. Palgrave Publication.

North, P. (1990.) *Institutions, institutional change and economic performance*. Cambridge, UK: Cambridge University Press:

Perrow, C. (1961). Organizational prestige. *American Journal of Sociology, 66,* 335–341.

Ruef, M., & Scott, W. R. (1998). A multidimensional model of organizational legitimacy: hospital survival in changing institutional environments. *Administrative Science Quarterly, 43,* 877–904.

Salmi, A. (1995). *Institutionally changing business networks: An analysis of a Finnish company's operations in exporting to the Soviet Unions, Russia, and the Baltic States.* Doctoral dissertation. Helsinki School of Economics and Business Administration.

Scott, W. R. (1995). *Institutions and organizations.* Thousands Oaks

Shrum, W., & Wuthnow, R. (1988). 'Reputational status of organizations in technical systems'. *American Journal of Sociology*, *93*, 882–912.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571 – 611.

Sydow, J., & Windeler, A. (1998). Organizing and evaluating inter firm networks: A structurationist perspective on network processes and effectiveness. *Organizational Science*, *9*(3), 265-284.

Vanhonacker, W. (2000). A better way to crack China. *Harvard Business Review*, July-August, 130-136.

Watts, C. (2003). http://www/ic.mr.icfai.org/casestudies/catalgoue/Business%20Strategy2/Motorola.ht.

Webster, L. (2002). *Motorola University Scores High Grades in China*. Asia Times Online Co. Ltd.

Wei, Y., & Bello, D. (2004). *Performance outcomes for business marketers in China: The moderating role of legitimacy behaviors.* ISBM Conference Workshop.

Wilkinson, I., & Young, L. (2000). On cooperating: Firms, relations and networks. *Journal of Business Research*, *55*(2), 123-132.

Zeng, P-Y. (2007). http://news.xinhuanet.com/english/2007-09/12/content_6711568.htm

KEY TERMS

3G (Third-Generation) Technology: This is usually used to describe the next wave of mobile communications. While analog technology represents the first wave of mobile communications technology and digital represents the second (2G), third-generation technology will provide more bandwidth for mobile devices.

Industrial Networks: In industrial networks, firms are linked together by their performance of industrial activities, employing or consuming various types of resources to produce other resources. These linkages create interdependence between firms, and the prior experience of interactions is important considerations because it limits the abilities of these firms to act independently.

Network Legitimacy: The quest for network legitimacy is motivated by the fundamental need of all firms to garner support from key network stakeholders, in order to survive and prosper. Crucial to their success is their ability to match their complimentary assets and targeted legitimacy orientations with these stakeholders.

Network Positions: Network position refers to the role a firm plays in the network and how it is linked directly and indirectly to other firms in the network.

Network Structure or Structuredness: This describes the dependencies of relationships between firms in the network and the resulting existence of an aggregate structure we characterize as a network.

Organizational Reputation: Reputation is describe as a core intangible resource and represents the affective or emotional evaluation of a firm as part of the social construction process, at a point in time, and over time.

Relationships: Relationship refers to conditions or character due to being related. Relationships arise between firms because of the interdependence of outcomes.

TD-SCDMA (Time Division Synchronous Code Division Multiple Access): This is China's home-grown third-generation mobile wireless technology. It competes with the US CDMA2000 system backed by Qualcomm Inc, and Europe WCDMA backed by Ericsson and Nokia. Development started in the late 1990s and is certified by the ITU (International Telecommunications Union) as a 3G (third-generation) standard in 2000.

Chapter XVII
Preliminary Knowledge Management Implementation in the Teleco Industry

Chin Wei Chong
Multimedia University, Malaysia

Siong Choy Chong
Putra International College, Malaysia

ABSTRACT

This chapter aims to create a unified model capturing and generalising the different arrays of preliminary knowledge management (KM) implementation success factors in the telecommunication industry based on the studies conducted on this sector in Malaysia. The literature and empirical evidence suggest that to become the leading global market players in the new knowledge society, telecommunication organizations are required to have the integration of an effective KM process which consists of construction, embodiment, and deployment. These processes must be supported by five preliminary success factors (clear business strategy; flexible organizational culture; a committed KM Team; and effective implementation of K Audit and K Map) and effective KM strategies, that is, a knowledge sharing culture and a human network, leadership, a wide range of new technological opportunities and increased commitment to measurement.

INTRODUCTION

The era of industrialisation and information age has enabled the telecommunication industry to expand into diversified functions to support the growth of technological advancement for better services demanded by any nation (Yusof, 1998). In its role as a key driver for electronic commerce and a facilitator to the diffusion of Internet, the telecommunication industry has transformed itself as well as economies worldwide (Insight Reports, 2005). The Insight Reports further reiterates that the importance of telecommunication industry is reflected in its growing share of world output, increasing over the last twenty five years from 1.6 percent of an estimated US$20 trillion in 1975 to 2.9 percent of an estimated US$40 trillion in global gross domestic product in 2000. The pace of growth has quickened in the last decade of the twentieth century and it has become a period of unprecedented growth for the telecommunication industry.

An example of this is the Malaysian telecommunication industry, wherein, the industry is expanding rapidly with the introduction of the latest services and equipment. For example, the use of broadband services was only at 0.85 percent in 2004 but its usage has increased to 5 percent in 2006 and is targeted to increase to 10 percent by 2008 (Lim, 2004). This development has become a catalyst for the growth of the nation's commercial and industrial sectors. The integration of the telecommunication and computer industries also resulted in the rapid growth of sophisticated technology which ushers in a new Information Technology-based century. The prominence of this sector is therefore not surprising, looking at its contributions as a tool of technological support for the nation's development in line with the development of the knowledge economy in Malaysia (Yusof, 1998). The Malaysian Government has developed policies to channel the telecommunication industry to foster a creation of information rich and intelligent nation via supreme telecommunication facilities and network. Due to the fast-paced developments in this industry both by pull and push factors, this sector is expected to move in a viable direction and thus bring in a better living scenario to the communities.

However, in this new millennium, this industry has to face an increasing level of unpredictability in its business environment and market competitiveness. The globalisation of business, shift from a production-based economy to a knowledge-based economy, growth of the information communications technology (ICT), the strive to become learning organizations and emergence of the need for knowledge workers are among the characteristics that have made knowledge management (KM) practice a must today across all types and levels of firms (Chong, 2006; Chong & Choi, 2005) and the telecommunication industry is not spared. All these changing environmental characteristics imply that the issues of more efficient and effective operation of the knowledge assets owned by the modern telecommunication organizations have become more critical than ever which contribute towards the continuing success of the firms in this competitive environment (Malhorta, 1998).

The importance of KM is also reflected by the strategic plans these telecommunication organizations have in order to become more competitive domestically and internationally. To achieve this, these firms need to continuously adapt and adopt the most effective technology and enhance their productivity. These organizations have therefore heightened their interest on KM practices to improve their efficiency and effectiveness in order to build and/or sustain their competitiveness. Given this backdrop, it is of paramount importance that the industry's KM practices is surveyed so that steps can be taken to improve and enhance the performance of these organizations in order to help the nation in realising its vision of becoming a developed nation by the year 2020.

PROBLEM STATEMENTS AND SIGNIFICANCE OF STUDY

Notwithstanding the importance and benefits of KM which have been widely documented (Nonaka, 1998; Sanchez et al., 2000; Teece, 2000; Tiwana, 2000), very few theoretical and empirical evidences are available with respect to the telecommunication industry's readiness to adopt KM (Chong et al., 2006; 2007; in press). Hence, KM implementation in the telecommunication industry appears to be an interesting area to be explored as these organizations struggle to adapt quickly, respond faster, and proactively shape their industries in this global yet turbulent environment.

Thus far, only four studies have investigated KM implementation among the telecommunication organizations (Chong & Yeow, 2005; Chong et al., 2006; 2007; in press). Chong and colleagues are probably among the first researchers to attempt a study on preliminary KM implementation, specifically the preliminary factors affecting the successful implementation of KM among the telecommunication organizations, particularly in Malaysia. However, different variables were investigated and therefore, a single, unified model of preliminary KM implementation is yet to emerge. In addition, while prior research have indicated a myriad of critical success factors of KM implementation (Chong, 2006; Chong & Choi, 2005), they have failed to determine which success factors are to be prioritised before a full-fledged KM implementation can take place. Some preliminary factors were proposed by different researchers (e.g. Nesbitt, 2002; Tiwana, 2000) but they were theoretical in nature and these factors have been looked upon in isolation. In other words, no attempts have been made to integrate them into a single framework. As such, in order to pursue KM implementation systematically, there is a need to identify, combine and empirically test all the preliminary KM success factors in one setting as the results would bear meaningful implications on the telecommunication organizations.

This research therefore aims to create a unified model capturing the different arrays of preliminary KM implementation success factors in the telecommunication industry. The outcomes of this study are of utmost importance both for research and practice. It serves as a guideline for the telecommunication organizations to undergo a self check and to help them rethink and reposition themselves in light of these findings. This helps to ensure that essential preliminary KM success factors are covered when the telecommunication organizations initiate KM implementation.

From the research perspective, it provides insights into what are the overall perception of KM and the key factors affecting successful implementation of KM among the telecommunication organizations. The resulting relationship will reinforce the findings of numerous studies about the importance of these factors to successful KM implementation. In short, it contributes to KM research by summarizing what has been done and identifying new areas that need to be explored in future research.

A UNIFIED FRAMEWORK OF PRELIMINARY KM IMPLEMENTATION

From the literature, it appears that KM is a systematic management of organizational knowledge, which involves the process of creating, gathering, storing, organising, diffusing, using, exploitation and assessment of knowledge of all kinds for creating business value and generating a competitive advantage (Chong, 2006). It often encompasses identifying and mapping intellectual assets within the organization, generating new knowledge for competitive advantage within the organization, making vast amount of corporate information accessible, and sharing of best practices through using technological tools as enablers. Based on these reviews, this study defines KM as the process through which organizations generate value from their intellectual and knowledge-based assets which involves codifying what employees, partners, and customers know and sharing that information among employees and departments in an effort to devise best practices.

Guided by the aforementioned definition of KM, most of the frameworks suggested by previous researchers focused mainly on the KM processes of creation, manifestation, use and transfer of knowledge (Coukos, 2001; Wong & Aspinwall, 2003). These are deemed cyclical process consisting of technology-centred strategy to manage explicit knowledge and people-centred strategy to manage tacit knowledge. These processes are most frequently associated with three types of activities: construction, embodiment and deployment (Coukos, 2001).

Coukos (2001) refers construction to the knowledge creation process which involves continuous, self-transcending activities such as development, discovery and capture of knowledge. The ultimate aim is to arrive at original, novel and useful ideas and solutions. The aim of the embodiment process, which involves the sub-processes of storing, categorising and mapping of knowledge, is to put organizational knowledge into a form that makes it accessible to those who need it. Its importance is reflected by the convenience of collecting and storing knowledge in an easily accessible and retrievable format. Deployment, or knowledge transfer, is seen as conveyance of knowledge from one place, person or ownership to another. It refers to the mechanisms used to make the repository content accessible.

At the same time, a considerable number of KM studies have also proposed several key variables for successful implementation of KM, coined as critical success factors (Chong & Choi, 2005, McAdam & Partkinson, 2003; Wong, 2005). These researchers have attempted to provide a comprehensive list of common success factors or key strategies in KM implementation. However, there are limited attempts in the identification of what constitutes preliminary success factors that need to be prioritised for KM implementation that need to be prioritised. Such determination is of paramount importance as it provides organizations with guidelines on which factors to concentrate on before launching a full-scale KM program.

Insofar, only two studies have identified preliminary KM success factors. Tiwana (2000) proposed four preliminary KM success factors: (1) KM and business strategy alignment; (2) knowledge map development (K Map); (3) knowledge assets audit (K Audit); and (4) KM team design. Nesbitt (2002) proposed business goals, K Audit, K Map and flexible organization creation. Some similarities and differences have been observed on the factors proposed but they have yet to be empirically proven.

The mainstream literature seems to arrive at the same conclusion that the efforts to link KM programs to business strategy have become a vital source of competitiveness for all organizations (American Productivity and Quality Centre, 1999; Tiwana, 2000). Tiwana (2000) rightly remarks that knowledge drives strategy and strategy drives knowledge management. As such, without a clear, articulated link between KM and business strategy, even the world's best KM program will deliver zilch (Levinson, 2005). Similarly, the effects of organizational structure towards the success of KM implementation have been well documented (Beijerse, 2000; Chong, 2006; Nonaka & Takeuchi, 1995). The literature dictates that an open and decentralised organizational structure best fit the organizational knowledge endeavours.

By the same token, the formation of a KM Team is also critical to the successful implementation of KM processes. This is because the major responsibility of the team is to build, implement, focus and deploy the KM systems. Similarly, it is widely acknowledged that knowledge of knowledge assets is critical to the proper planning of a KM program and/or systems. It provides a rich source of information about the strengths of the organization that must be exploited for its full value to be realised by the owners (Teece, 2000). The audit process involves examining all the intangible assets, which include organizational rituals, processes, structures, communities and people and determining the monetary value of these assets. The areas that hold the most potential for future growth and strategic advantage will be selected and invested by organizations.

K Map, on the other hand, provides a snapshot of where an organization is at any given time relative to its competitors (Tiwana, 2000). K Map is needed as it helps to describe how to find, what to find and where to find useful knowledge within the organization.

Some researchers proposed several key KM strategies instead of success factors. KM strategies work simultaneously to streamline and enhance the capture, flow and transfer of an organization's data, information, and knowledge for the purpose of delivering it to individuals and groups engaged in accomplishing specific tasks (Coukos, 2001; O'Dell & Grayson, 1999; Sveiby, 2000). Among the strategies that have received most attention are technology, culture, leadership and measurement. Technology has long been regarded as a key enabler for implementing a successful KM program (Allee, 2001; Chong, 2006). With regards to KM, technology consists of infrastructure of tools, systems, platforms, and automated solutions that are centralised to enhance the development, application and distribution of organizational knowledge (Coukos, 2001). Culture refers to the general knowledge sharing climate of an organization as related to an integrated pattern of human behaviours – including thoughts, speeches, actions, and artefacts (Coukos, 2001). It has been widely acknowledged that people and culture are at the heart of creating a successful knowledge-based organization (Chong & Choi, 2005).

Top management support and leadership is another important strategy towards successful KM implementation (Chong, 2006; Coukos, 2001; Jones et al., 2003). This is because organizational leaders have the ability to align KM behaviours with organization strategies, identify opportunities, promote the values of KM, communicate the best strategies, facilitate the evolution of learning organizations and provide metrics for assessing the impact of knowledge (Coukos, 2001). Finally, measurement has been referred to as the assessment methods of KM and its relationships with organizational performance. However, Ahmed et al. (1999) found that many companies fail to measure their KM efforts due to reasons such as failure to operationally define performance, failure to relate performance to the process, failure to define the boundaries of the process, and misunderstood or misused measures.

In short, a comprehensive model was derived at based on the review of literature. This paper proposes three main elements characterising preliminary KM practices, namely three main KM processes (construction, embodiment and deployment), five preliminary success factors (business strategy, organizational structure, KM team, K Audit and K Map) and four KM strategies (culture, leadership, measurement and technology) that have garnered impressive empirical and/or theoretical support. The next section and sub-sections provide empirical results of the KM processes, preliminary KM success factors and KM strategies based on the studies conducted on the telecommunication industry in Malaysia.

AN INVESTIGATION FROM MALAYSIAN TELECOMMUNICATION INDUSTRY

To investigate the relevance of this framework and how it is applied, a study on the Malaysian telecommunication industry has been conducted. The population frame of the telecommunication companies comprises of three categories, i.e. line providers, manufacturers of mobile phones and application service providers. A non-probability sampling method, i.e. convenience sampling technique was adopted since it is not easy to employ a random sampling method, looking at the complex nature of the industry. Forty companies were identified and contacted. Twenty-two companies agreed to participate in the survey. Forty copies of questionnaires were sent to middle managers from each of the twenty-two companies. The middle managers were chosen to participate in this survey due to the important role played by them in successful KM implementation (Chong, 2006; Nonaka & Takeuchi, 1995; Salleh & Goh, 2002). Out of 880 questionnaires distributed, 289 questionnaires were completely filled, yielding a response rate of 33 percent.

The descriptive statistics revealed that 29.8 percent of the respondents indicated that they have been implementing a full scale KM program in their organizations. Approximately 37 percent of the respondents indicated that they already have one or more pilot applications on KM while 31.8 percent of them were at the planning and evaluation stage. The results confirmed previous findings (Chong & Yeow, 2005; Chong et al., 2006a) that most of the Malaysian telecommunication organizations are still at the beginning stage of KM implementation. As this study attempts to test the pre-requisite factors prior to KM implementation, the nature of the telecommunication industry is deemed to be appropriate.

Data was collected using a set of self-reporting questionnaire. The questionnaire comprises of two sections, the first section consists of questions on organizational demographic characteristics while the second section comprises of fifty-nine questions measuring the respondents' perceived importance and degree of implementation of KM processes (15 items) preliminary success factors (24 items) and KM strategies (20 items) using a 5-point Likert scale from 1 (not important/not implemented) to 5 (very important/extensively implemented). The questions were developed based on the extensive review of literature (Coukos, 2001; Tiwana, 2000).

The Cronbach alpha's score between 0.9667 and 0.9744 for all the variables indicate that the instrument is highly reliable. Face validity and construct validity were also achieved based on the pilot testing undertaken and the results of factor analyses explained 69.751 percent to 87.044 percent of the variances (Chong et al., 2006; 2007; in press) (Table 1). Generally, all the variables have eigenvalues of 1 or greater and none of the items were dropped since they have coefficients of over 0.50. As expected, all the attributes are found to be appropriately grouped under the sub-groups of KM processes, preliminary success factors and strategies, confirming that the results of this study are in congruence with theory and practice.

The following section presents the comparison

Table 1. Factor analysis results of KM processes' items

	Eigenvalue		Variance (%)		Cronbach's Alpha	
	FL (PI)	**FL (AI)**	**FL (PI)**	**FL (AI)**	**FL (PI)**	**FL (AI)**
Process 1: Construction	*3.755*	*4.068*	*75.096*	*81.369*	0.9170	0.9427
Process 2: Embodiment	*3.591*	*3.859*	*71.814*	*77.186*	0.9017	0.9258
Process 3: Deployment	*3.488*	*4.134*	*69.751*	*82.682*	0.8911	0.9472
Factor 1: Business Strategy	*3.60*	*3.876*	*72.004*	*77.524*	0.9015	0.9260
Factor 2: Organizational Structure	*2.836*	*3.254*	*70.903*	*81.357*	0.8621	0.9231
Factor 3: KM team	*3.385*	*4.154*	*67.693*	*83.080*	0.8799	0.9489
Factor 4: K Audit	*3.824*	*4.126*	*76.478*	*82.526*	0.9228	0.9469
Factor 5: K Map	*3.977*	*4.261*	*79.548*	*85.224*	0.9356	0.9566
Strategy 1: Technology	*3.937*	*4.258*	*78.730*	*85.169*	0.9320	0.9560
Strategy 2: Culture	*3.672*	*3.839*	*73.439*	*76.778*	0.9095	0.9242
Strategy 3: Leadership	*4.068*	*4.258*	*81.356*	*85.167*	0.9424	0.9555
Strategy 4: Measurement	*4.101*	*4.352*	*82.023*	*87.044*	0.9452	0.9627

FL (PI) – Factor loading for perceived importance;
FL (AI) – Factor loading for actual implementation

results and subsequently the recommendations to enhance KM implementation in the telecommunication industry.

FINDINGS, IMPLICATIONS & RECOMMENDATIONS

Table 2 presents the results of the paired t-test analysis between all the factors under KM processes, KM preliminary success factors and KM strategies. The results reveal that there are significant differences between all the factors perceived as important and their levels of implementation.

Table 2. Comparison of the perceived importance and implementation of KM preliminary success factors, KM strategies and KM processes

	Mean Difference	*t*-value	*Sig. value*
KM Processes			
1. Construction	.97	14.67*	0.000
2. Embodiment	.98	15.42*	0.000
3. Deployment	.93	14.65*	0.000
Average mean difference	**.96**		
Preliminary Success Factors			
1. Business Strategy	.89	13.93*	0.000
2. Organizational Structure	.96	13.13*	0.000
3. KM team	.96	13.40*	0.000
4. K Audit	1.04	15.15*	0.000
5. K Map	.99	14.58*	0.000
Average mean difference	**.97**		
KM Strategies			
1. Technology	.81	12.51*	0.000
2. Culture	.86	14.19*	0.000
3. Leadership	.93	14.03*	0.000
4. Measurement	1.00	14.31*	0.000
Average mean difference	**.90**		

In summary, the Malaysian telecommunication organizations in general are aware of the importance of the KM elements but are not fully prepared with their implementation. This is in view of the fact that most of the telecommunication organizations have just begun to implement KM (Chong & Yeow, 2005) and therefore, they are not aware of the whole spectrum of KM implementation. This is evident from the descriptive statistics which reveal that only 29.8 percent of the respondents have implemented a full-scale KM program while the remaining are either in the pilot or planning stage. There appears to be a lack of understanding among the telecommunication companies of what is needed for a KM program that might have affected their focus and thus their level of implementation. Lack of time, manpower and cost have also been singled out as challenges facing the telecommunication firms (Chong et al., 2006; 2007; in press).

This finding is not uncommon as it is in line with many prior research findings conducted in different sectors in Malaysia and also in the United States. Such consistent findings lead one to safely conclude that lack of guidance of proper KM implementation has caused the organizations to overlook some of the important aspects. This research is thus timely in order to guide the telecommunication organizations toward proper initiation of KM initiatives. The following subsections review the findings of KM processes, preliminary success factors and strategies and the recommendations drawn from the findings.

KM Process — Construction

A higher significant difference was found between the level of importance and implementation of the construction process. This is not surprising as it was discovered that the telecommunication firms have invested less efforts in examining and generating new knowledge (Chong et al., in press). In order to enhance the construction process, the telecommunication organizations should consider investing more efforts in generating new knowledge through different approaches such as hiring new employees, forming partnerships with other organizations, outsourcing some of the functional areas, engaging external consultants, seeking help from customers or professional associations, attending external as well as in-house training. They ought to upgrade themselves with sophisticated and effective methods in searching for new information.

This, however, does not mean that the telecommunication organizations should create new knowledge from scratch. They may configure and recombine existing pieces of knowledge, along with the strategy of imitation, replication and substitution. The key here is to focus on its capabilities while limiting its shortcomings. The knowledge gained from external sources will allow the telecommunication firms to get better perspective of their knowledge bases which may include knowledge from outside (Bhatt, 2000). Since knowledge gets obsolete fast enough, its relevance at any given time changes, as do the skills of employees. An organizational-wide culture of caring, sharing and training is thus needed to ensure the effectiveness of the knowledge construction process. More importantly, the construction of knowledge must be a consistent and continuous effort.

KM Process — Embodiment

Table 2 shows that higher significant difference was found between the degree of importance and the implementation level of the embodiment process among the telecommunication organizations. This is not surprising since many of them are in the early stage of KM implementation; they need time and manpower to codify their existing tacit knowledge into documentation. In order to narrow the gap, employees must be given proper training and guidance to assist them in codifying their knowledge into a form of documentation. Employees have to be made aware of the importance of the embodiment process as it helps to keep and retain the expertise and knowledge within the telecommunication organizations. Although they might not be able to reuse past knowledge in the current era, these documentations serve as a good reference and useful guidelines to them. Since the embodiment process requires time, the telecommunication organizations should not delay the implementation of this process.

KM Process — Deployment

While lower significant differences are found between the level of importance and the implementation level, the significant difference that exists suggest that the implementation level could be enhanced. The telecommunication organizations should be aware that knowledge transfer is critical to organizational learning and success as it impacts on organizational productivity and quality which in turn leads to organizational competitiveness. If tangible assets depreciate in value when they are used, knowledge grows when used and depreciates when not being used (Syed Ikhsan & Rowland, 2004).

While tacit knowledge can be embedded in procedures or presented in documents and databases with reasonable accuracy, tacit knowledge generally requires extensive personal contacts. The telecommunication organizations have to start finding effective ways to encourage people talking and listening to each other and get them to work based on the knowledge acquired. The management plays an important role to encourage informal discussions, meetings and informal gatherings among their respective departments to exchange and share their knowledge. Shadowing and joint problem solving are probably two of the best practices for transferring important knowledge by pairing more experienced with new organizational members. In addition, proper infrastructure, including technological tools, has to be in place to ensure successful knowledge transfer process. The speed resulted from knowledge transfer may lead to speedier deployment of knowledge that benefits most organizational members in terms of decision making. However, the organizations must be careful that only accurate data, information and knowledge are transferred through the systems.

KM Preliminary Success Factor — Business Strategy

Although a lower gap is identified, its significance level suggests that there are more to be done by the telecommunication firms in linking their KM programs with their business strategies. Since many of the telecommunication firms are in their infancy stage of KM implementation, they cannot see how a clear and well planned KM strategy is important for the organizations to achieve their business objectives. To bridge this gap, the management of these firms should understand the vital role they played in ensuring the consistency between their KM strategies and business goals. In other words, the major impact of knowledge has to be considered when an organization formulates its corporate strategy. The management needs to ensure that their knowledge strategy and knowledge program is consistent with corporate ambitions, and that the techniques, technologies, resources, roles, skills and culture are aligned with and support the business objectives. At the same time, efforts have to be taken to formulate an appropriate and full KM plan which can holistically integrate the knowledge capabilities of all departments to create value for the telecommunication organizations. Notwithstanding the roles played by the management, the respective human resource departments of the telecommunication organizations play important roles to educate and disseminate information to each and every employee on the purpose and benefits of KM.

KM Preliminary Success Factor — Organizational Structure

While the respondents rated an open and decentralised organizational structure as important in facilitating KM activities, the empirical evidence implies that the firms do not have an accurate understanding of the organizational structure that best fit a knowledge-intensive organization. This suggests that improvements to the existing organizational structure of the telecommunication organizations are possible in order to reap successful KM implementation.

The telecommunication organizations must realise that an open organizational environment based on flexibility, variation and renewal must be encouraged. As such, the existing structural characteristics of the organizations have to be changed. Beijerse (2000) rightly suggests that knowledge-intensive organizations have to be ideally flat, with short communication lines between employees and management and that consultation is done when setting up the structure of product-market combinations and/or projects that are centralised and require coordination. Nonaka and Takeuchi (1995) propose a new management process called middle-up down and bottom-up management models, and it is the most fitting model for bringing about organizational knowledge creation. Such decentralised structures tend to respond more rapidly to change and will help support and facilitate the learning process within the telecommunication organizations to achieve their goals in the most efficient way. To begin with, the telecommunication organizations can achieve this in several ways such as gathering experts from different departments when forming project teams, designing employees' work stations in a way to ease the flow of ideas across departments, and/or providing venues for employees to communicate informally.

KM Preliminary Success Factor — KM Team

The significant difference found between the degree of importance and implementation levels of KM Team shown in Table 2 indicate that the telecommunication organizations could further improve on this effort. While many of the organizational members of these telecommunication firms work in teams, their activities are limited to achieving business performance rather than KM initiatives. It is acknowledged that good leadership ensures the success of effective KM teams. The telecommunication organizations could establish a KM leadership position such as the Chief Knowledge Officer (CKO) or to second a senior management member to the position if finance is a major concern. Having a supportive leadership is imperative to the success of KM implementation as the leader performs many valuable KM activities with the KM team such as to design, implement and oversee a firm's knowledge infrastructure, measure and manage the value of knowledge and advocate knowledge discovery and use (Jones et al., 2003). It is also suggested that the KM team comprises of employees from various departments so that they can collaborate with others to implement KM activities. To enhance the success of the team, it is important for the influencers across the organization to be identified and made members of the KM teams. These influencers consist of employees who are well respected by their peers and whose opinions are highly recognised in their organizations. This is because when the members have good things to say about their KM efforts; their positive attitudes will go a long way towards convincing others on the merits of KM.

KM Preliminary Success Factor — K Audit and K Map

Despite the importance of K Audit and K Map as reported in the literature, these factors scored average mean differences as shown in Table 2. It was observed that both factors fall short of implementation mainly due to resources constraints and lack of expertise (Chong et al., in press). This is not uncommon as it was reported that many firms do not know where to find their existing knowledge (Davenport & Prusak, 1998) and that K Audit requires time and experts' help to be accomplished. Similarly, the K Map process is time consuming and might be very costly especially during the start up phase.

The telecommunication organizations should be aware of the importance of K Audit and K Map as they are critical activities that facilitate successful implementation of KM. In fact, both activities are needed before the firms could institutionalise full-fledged KM initiatives. At the fundamental level, they should view costs associated to these activities as investment in which the payoff is greatest if KM can be successfully implemented. If lack of expertise is the case for many organizations, the firms should consider engaging in a group of experts; possibly external consultants since the majority of them are in the beginning stage of KM implementation. These experts will be able to help review the types of knowledge that the organizations have, figure out areas of knowledge that they are lacking and identify well-defined programs and strategies in order to close the knowledge gap to improve competitiveness. This gap should then be rectified through excellent execution of core capabilities. Even the consultants can help to develop a proper K Map for the organizations indicating where knowledge, expertise and experiences resides (in people, documents, processes) and which knowledge needs to be shared with whom, when how and why.

At the same time, the human resources of these organizations need to be trained so that eventually internal teams can be formed to coordinate, initiate and manage the implementation process. Another important aspect to be considered by the telecommunication organizations is to establish proper mechanisms to solicit feedback from employees, customers, suppliers and other stakeholders. The resulting information gathered is important especially in the telecommunication industry where the ability to understand market needs is especially critical to fulfil the needs and requirements of customers. The organizations should keep in mind that both K Audit and K Map have to be a continuous process and that the results need to be documented in a consistent capability framework in order to allow for comparison over time. The continuous process allows the telecommunication organizations to be well aware of what knowledge assets they have and know their strengths and limitations in the quest of becoming market leader to be continuously in search for competitiveness in the marketplace.

KM Strategies — Culture

While culture has been identified as one of the most critical element to the successful implementation of KM as do the middle managers when they rated on the importance of this element, the findings imply that the cultural practices in the telecommunication organizations fail to support their KM endeavours.

The telecommunication organizations should understand that creating and sharing knowledge are intangible activities that cannot be forced. The top management has an important role to play to create and preserve proper organizational climate conducive to knowledge sharing. They ought to realise the importance of people as the foundation of KM and therefore, they have to be motivated to access and share information and to convert that information into knowledge (Brand, 1998). As such, an innovative culture has to be established in which individuals are constantly encouraged to generate new ideas, knowledge and solutions. Employees must be allowed to commit mistakes and to learn from failures. On top of that, employees must be empowered in their jobs because through empowerment, employees take extra responsibilities to solve organizational problems by learning new skills in their jobs, leading them to being more knowledgeable and competent. Employee involvement is another area to enhance KM activities within the telecommunication organizations. To achieve this, a proper reward and knowledge-based performance management systems must be institutionalised. Besides, the human resource departments can play an important role by creating a supportive environment for KM programs through regular staff meetings, creation of office space for staff to meet informally, sharing of KM success stories in organizational newsletters and so on.

KM Strategies — Leadership

Table 2 provides empirical evidence that the telecommunication organizations have not been very effective in promoting KM, judging from the higher and significant level of differences. This is not surprising, given the fact that KM is still new to the telecommunication firms and many of them lack capable and experienced leaders in managing KM effectively.

As stressed in various parts of this article, top management has been instrumental in ensuring KM implementation to be a success. They must render their full support towards KM implementation. This involves developing a strategic vision in the management of the firms' intangible assets. In addition, top management has the ability to show its support by creating a KM leadership position such as CKO.

While the organizations can second a senior management member to the position, it is strongly encouraged that the job incumbent is hired from outside the organization because he or she is likely to possess knowledge that the individuals within the firm do not know. The organizations should understand that the CKO's key function is their understanding of the organization and its business drivers, combined with the ability to take a holistic view of the company and to understand the mix of hard and soft skills necessary to create, sustain and utilise the firm's knowledge base (Jones et al., 2003). Their main responsibility is to manage the knowledge of an organization by converting knowledge into profit by leveraging the organization's intellectual assets. The CKO can work with the KM Team to facilitate knowledge sharing and increase the effective use of organizational memory by working with other employees throughout the organization to codify and institutionalise new knowledge. By having a CKO with a clear sense of the full scope and depth of responsibilities in KM, it will enable the entire knowledge chain to be supported and hence allow successful implementation of KM in the telecommunication organizations.

KM Strategies — Technology

Technology has long been regarded as a key enabler for implementing a successful KM program. It is expected that most of the telecommunication organizations are equipped with advanced technology tools and methods due to the nature of their businesses (Chong et al., 2007); however, it has been found that the telecommunication organizations have not been very effective in utilising technology in their KM efforts. This is not surprising as the organizations have just started to implement KM and therefore, information systems and/or technology cannot play a significant role. Many of the organizations have just begun to develop their technology infrastructure.

To enhance the implementation level of technology in the KM efforts undertaken by the telecommunication organizations, it is suggested that organizations that create knowledge-based systems are more successful in their KM endeavours as such systems support the creation, harvesting, assimilation and leveraging of knowledge (Soliman & Spooner, 2000). Many tools are available to support these knowledge functions, such as relational databases, text and document search engines, data warehouses and data sharing tools. Groupware is another important tool which includes the ability to send and receive e-mail, share personal calendars, hold computer conferencing, and workflow management. Examples are Lotus Notes, e-mail, intranet, internet, electronic conferencing tools and many others. These are some of the applications that could be used by the telecommunication organizations.

It is not necessary for the organizations to develop their KM systems from scratch because the best systems is one that fit their requirements and that they could build upon what they already had. Important factors that need to be considered in the development of a KM system include simplicity of technology, ease of use, suitability to users' needs, relevancy of knowledge content, and standardisation of a knowledge structure or ontology. More importantly, management needs to ensure their employees know how to use these tools by sending them for training.

KM Strategies — Measurement

The importance of a systematic measurement system towards the success of KM has been well documented. However, a larger significant difference (Table 2) implies that the telecommunication firms do not have an adequate understanding of this strategy. In fact measurement has been identified as the least implemented factor (Chong et al., 2007). It can thus be concluded that the firms are not aware of the implications of performance measurement due to their infancy stage of KM implementation. Further, the development of a comprehensive performance measurement system is yet to exist, and therefore, the firms lack guidance when it comes to developing one.

The telecommunication firms should understand the fact that KM evaluation is incomplete by depending on financial measures alone. They should develop metrics to allow the business to measure its impacts, provide room for improvement and to provide a robust basis for resource allocation (Duffy, 2000). A combination of measurement approach, which includes quantitative methods, qualitative assessment, performance review and benchmarking, is needed due to the tacit nature of knowledge and its dynamism.

While it is understood that a comprehensive measurement model is yet to be established, the telecommunication organizations could try different approaches such as Bohn's (1994) eight-stage of knowledge growth. The American Productivity and Quality Centre (2001) have also proposed a formal measurement on the effectiveness of knowledge-intensiveness business processes. The balanced scorecard technique (Kaplan & Norton, 1992) provides another avenue for the telecommunication firms to measure their KM initiatives against business success.

FUTURE RESEARCH

The results of the current study indicate that there is a need to continue investigating and examining the preliminary success factors, strategies and processes related to KM in telecommunication organizations. The knowledge era calls for different behaviour – behaviour that is more experimental and risk-taking. Telecommunication organizations will need to become more proactive in order to take advantage of opportunities and compete strategically for limited resources, especially in this competitive environment. A cross-cultural research based upon the proposed model is possible, in view of cultural differences that might affect management philosophy. A cross-industry survey is also warranted to compare the findings between the telecommunication industry vis-à-vis other sectors. In addition, it is interesting to determine if there are any new factors influencing successful KM implementation that has not been reported in this study. Further to this, as the intercorrelation between the three variables is hypothesised, a study on the linkage between the KM processes, success factors and strategies and possibly with organizational demographics (i.e. number of employees, size of organizations and so on) is also necessary. These will undoubtedly allow the survey results to be more conclusive and accurate, which will subsequently allow for meaningful generalisations of the results to be made.

CONCLUSION

This study extends knowledge in KM, especially concerning the early stage of KM implementation among the telecommunication industry. This study serves as a foundation for building a cumulative tradition of research, particularly in an industry which is in its early stage of KM implementation.

KM, an emerging private sector philosophy, provides a framework for addressing organizational and competitive challenges as it requires the integration of an effective KM processes with the success factors and strategies. Because KM deals with cultural, strategic, and technological issues, it is imperative for the telecommunication organizations to realign its focus on the KM processes, success factors and strategies. The key to successful KM implementation lies in people and therefore the roles of top management, culture and infrastructure must be emphasised. By focusing on strategic planning and execution of the KM elements proposed in this study, the telecommunication organizations can be assured of a results-oriented KM practice where needs are met, opportunities are approached, and organizational performance is improved.

If the telecommunication organizations wish to achieve their potential in becoming leading global market players in the new knowledge society and the primary location where knowledge is conceived and designed, they must address the challenges presented in this study. Nevertheless, this work advances the much needed baby step in in achieving these ends.

REFERENCES

Ahmed, P. K., Lim, K. K., & Zairi, M. (1999). Measurement practice for knowledge management. *Journal of Workplace Learning: Employee Counseling Today, 11*(8), 304-311.

Allee, V. (2001). *12 principles of knowledge management.* American Society for Training and Development (ASTD).

American Productivity and Quality Centre (APQC) (2001). *Successfully implementing knowledge management.* APQC: Consortium Learning Forum Best Practice Report.

Beijerse, U. R. P. (2000). Knowledge management in small and medium sized companies: Knowledge management for entrepreneurs. *Journal of Knowledge Management, 4*(2), 162-179.

Bhatt, G. D. (2000). Organising knowledge in the knowledge development cycle. *Journal of Knowledge Management, 4*(1), 15-26.

Bohn, R. E. (1994). Measuring and managing technological knowledge. *Sloan Management Review, 36*(1), 61-73.

Brand, A. (1998). Knowledge management and innovation at 3M. *Journal of Knowledge Management, 2*(1), 17-22.

Buckman, R. (1999). Collaborative knowledge. *Human Resource Planning, 22*(1), 22-33.

Bukowitz, W., & Williams, R.L. (2000). *The knowledge management fieldbook*. Revised edition, London: Prentice-Hall.

Chong, C. W., Chong, S. C., & Yeow, P. H. P. (2006). KM implementation in Malaysian telecommunication industry: An empirical analysis. *Industrial Management & Data Systems, 106(*8), 1112-1132.

Chong, C. W., Chong, S. C., & Wong, K. Y. (2007). Implementation of KM strategies in the Malaysian telecommunication industry: An empirical analysis. *VINE: The Journal of Information and Knowledge Management Systems, 37*(4), 452-470.

Chong, C. W., Chong, S. C., & Wong, K. Y. (in press). Is the Malaysian telecommunication industry ready for KM implementation. *Journal of Knowledge Management.*

Chong, C. W., & Yeow, P. H. P. (2005). *An empirical study of perceived importance and actual implementation of knowledge management process in the Malaysian telecommunication industry.* Paper presented at the International Conference of Technology and Management: Challenges and Prospects. (pp. 182-192), Faculty of Business and Law, Multimedia University, Melaka.

Chong, S. C. (2006). KM implementation and its influence on performance: An empirical evidence from Malaysian Multimedia Super Corridor companies. *Journal of Information and Knowledge Management, 5*(1), 21-37.

Chong, S. C., & Choi, Y. S. (2005). Critical factors in the successful implementation of knowledge management. *Journal of Knowledge Management Practice, 5.* Retrieved February 13, 2006, from http://www.tlainc.com/articl90.htm.

Coukos, E. D. (2001). *Knowledge management: Processes and strategies used in research universities.* Unpublished doctoral dissertation, Florida Atlantic University, Boca Raton, FL.

Davenport, T., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know.* Boston, Massachusetts: Harvard Business School Press.

Demarest, M. (1997). Understanding knowledge management. *Long Range Planning, 30*(3), 374-384.

DeSanctis, G., & Gallupe, B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(12), 1589-1609.

Duffy, J. (2000). Knowledge management: To be or not to be? *Information Management Journal, 34*(1), 64-67.

Gumbley, H. (1998). Knowledge management. *Work Study, 47*(5), 175 -177.

Insight Reports (2005). *Telecommunications and VoIP in Asia, Oceania and the mideast: a market perspective on the major economies 2005-2011.* The insight research corporation, USA.

Jones, N. B., Herschel, R. T., & Moesel, D. D. (2003). Using knowledge champions to facilitate knowledge management. *Journal of Knowledge Management, 7*(1), 49-63

Kalsom, S., & Syed Noh, S. A. (June, 2006). *KM strategy for E-Government: An exploratory study of local authorities in Malaysia*. Paper presented at the Knowledge Management International Conference and Exhibition 2006 (257-265), Legend Hotel, Kuala Lumpur, Malaysia.

Kaplan. R. S., & Norton, D. P. (1992). The balanced scored measures that drive performance. *Harvard Business Review, 70*(1), 71-79.

Levinson, M. (2005). The ABCs of KM. *Knowledge Management Research Centre*. Retrieved January 22, 2008, from http://www.cio.com/research/knowledge/edit/kmabcs.html.

Lim, K. Y. (2004). *Communications in Malaysia*. Retrieved March 19, 2005, from http://en.wilipedia.org/wiki/Communications_in_Malaysia.

Malhorta, Y. (1998). Deciphering the KM hype, *Learning and Information Management, 21*(4), 58-60.

Moffett, S., McAdam, R., & Partkinson, S. (2003). An empirical analysis of knowledge management applications. *Journal of Knowledge Management, 7*(3), 6-26.

Nesbitt, K. (2002). *Designing a knowledge management system*. Retrieved March 19, 2005, from http://academic.edu:2081/products/faulknerlibrary/00018382.htm.

Nonaka, I. (1998). The knowledge-creating company. *Harvard Business Review on Knowledge Management*. Boston, MA: Harvard Business School Press.

Nonaka, I. (1991). The knowledge creating company. *Harvard Business Review,* November – December, 96.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

O'Dell, C., & Grayson, J. (1999). *If only we know what we know: The transfer of internal knowledge and best practice*. New York: Simon and Schuster.

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences.* San Francisco: W.H. Freeman & Co.

Salleh, Y., & Goh, W. K. (2002). Managing human resources toward achieving knowledge management. *Journal of Knowledge Management, 6*(5), 457-468.

Sanchez, P., Chaminade, C., & Olea, M. (2000). Management of intangibles: an attempt to build a theory, *Journal of Intellectual Capital, 1*(4), 312-327.

Silos, I. (1999). Employee involvement – a component of total quality management. *Production & Inventory Management Journal, 40*(1), 56-65.

Soliman, F., & Spooner, K. (2000). Strategies for implementation knowledge management: Role of human resource management. *Journal of Knowledge Management, 4*(4), 337-345.

Sveiby, K. E. (2000). *What is knowledge management.* Retrieved March 18, 2005, from http://www.sveiby.com.ay/KnowledgeManagement.html.

Syed-Ikhsan, S. O. S., & Rowland, F. (2004). Benchmarking knowledge management in a public organization in Malaysia. *Benchmarking: An International Journal, 11*(3), 238-266.

Teece, D. (2000). Strategies for managing knowledge assets: The role of firm structure and industrial context. *Long Range Planning, 33*(1), 35-54.

Tiwana, A. (2000). *The knowledge management toolkit: Practical techniques for building a knowledge management system*. 1st ed., Upper Saddle River, NJ: Prentice-Hall PTR.

Verespej, M. (1999). Knowledge management: System or culture? *Industry Week, 248,* 20-21.

Wong, K. Y. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial Management & Data Systems, 105*(3), 261-279.

Wong, K. Y., & Aspinwall, E. (2003). *Is knowledge management equivalent to information technology?* Paper presented at the Fourth European Conference on Knowledge Management (989-997), Oxford University, Oxford.

Yusof, A. M. (1998). *The future path of Malaysian telecommunication industry*. Retrieved March 18, 2005, from http://members.tripod.com/~azman97/essay7.html

KEY TERMS

Construction: It comprises a set of activities associated with the entry of new artefacts into the system and includes such activities as development, discovery, and capture.

Deployment: Knowledge transfer represents the mechanisms used to make repository content accessible.

Embodiment: Embodiment refers to translating data and information into symbols that others can understand.

K Audit: The knowledge of knowledge assets and is a rich source of information about the strengths of an organization.

K Map: K Map provides a snapshot of where an organization is at any given time relative to its competitors.

KM Team: A group of people whose responsibility is to build, implement, focus and deploy a KM program.

Knowledge Management: Process through which organizations generate value from their intellectual and knowledge-based assets which involves codifying what employees, partners, and customer know and sharing that information among employees and departments in an effort to devise best practices.

Chapter XVIII
Activity-Based Costing in the Portuguese Telecommunications Industry

Maria Major
ISCTE – Business School, Portugal

Trevor Hopper
University of Manchester, UK

ABSTRACT

This chapter examines an implementation of an Activity-based Costing (ABC) system in a Portuguese telecommunications firm called Marconi. It is argued that they changed its management accounting system due to efficiency and institutional pressures from its constituencies following the liberalisation of the Portuguese and European Union telecommunications market. The replacement of Marconi's "old" management cost accounting system by an ABC system popularised by the business mass media and consultants as a "progressive" management accounting tool, helped display expectations of dynamism and efficiency to important external institutions, such as regulators. Whilst it was used and welcomed by managers dealing with commercial matters, operational managers were more unconvinced and contributed to technical problems affecting its accuracy and operation. The divided reactions of the managers give support to both advocates of ABC and its critics who claim it is beset with problems of economic measurement, behavioural issues during implementation and operation, and questionable cost-benefit returns.

INTRODUCTION

The last decade has witnessed unprecedented changes in the European telecommunications industry consistent with international trends. Technological innovation, growing customer demand, and European Union (EU) and the World Trade Organization pressures to deregulate national telecommunication sectors have forced many EU countries' governments to restructure their national telecommunications sector. Portugal is one such example. Until the mid-1990's its telecommunications sector was not open to competition, and telecommunications companies[1] were operated as monopolies and state concessionaires. Most were production oriented and were frequently criticised for inefficiently spending public resources and neglecting market pressures.

Following EU policies introducing liberalisation in the mid-1990s the Portuguese telecommunications sector was drastically restructured by: creating a 'powerful' Portuguese telecommunications group (called 'Portugal Telecom Group'); the establishment of new (private) telecommunications firms; and the active monitoring of the market by the Portuguese telecommunications regulator – ICP[2]. The aims of this programme executed by the Portuguese government under EU supervision were to create adequate conditions for introducing full competition; to prepare the market for new operators; and to restructure old inefficient public operators.

This forced 'old' telecommunications operators to improve their management accounting systems (MAS), which were inadequate for the new demands of the regulator and managers regarding relevant data. Marconi, which provided long distance telecommunications services, was one of several companies that changed their MAS to an Activity-Based Costing (ABC) system during the late 1990's. Apart from efficiency reasons and a concern to control costs, the adoption of ABC was also associated with a desire to reassure the EU, the national regulator and its constituencies – competitors, customers and shareholders - that interconnection prices were cost oriented and costs were accurately calculated. The latter was particularly important as Marconi (and its parent, Portugal Telecom) had to interconnect new operators to their telecommunications network[3]. The adoption of ABC, regarded as a rational and modern cost accounting system, helped Marconi gain legitimacy from external constituents for its interconnection pricing.

The aims of this chapter are threefold: first, it examines the merits and pitfalls of ABC as debated by accounting researchers; secondly, it describes the management accounting (MA) changes, namely ABC, in a specific telecommunications company; and thirdly, it assesses how ABC operated. The chapter is structured thus: the next section examines criticisms of traditional MASs and how ABC purports to resolve them in a contemporary business environment. It also outlines the main features of ABC and points out that despite its popularity it has received considerable criticism and reservations. Then details of the company (Marconi) where our case study was conducted are provided, along with an account of how it was implemented ABC, how it was operated, and consideration of whether it generated effective data for decision-making and associated behavioural problems.

ACTIVITY-BASED COSTING

Although activity costing was referred to earlier by Staubus (1971) and Shillinglaw (1982) its popularity stemmed from publicity of its application in a few US companies in the mid 1980s in case studies published by Cooper and Kaplan. They alleged that traditional MA was inappropriate for a contemporary business environment; it often provided decision-makers with inaccurate cost information; it lagged behind manufacturing technology changes; and it no longer provided relevant information for managers (Johnson and Kaplan, 1991; Innes and Mitchell, 1996). Scapens (1994) suggested that MA delayed technological changes because accounting practices form routines and rules that once institutionalised become hard to change. It was widely claimed that current cost accounting systems had lost their relevance by the 1980s (Kaplan, 1984; Johnson and Kaplan, 1991) and MA should return to basics, "working closely with design and process engineers, operations managers, and product and business managers" (Johnson and Kaplan, 1991: 261). Johnson and Kaplan (1991) claimed MA's lost relevance stemmed from allocating indirect costs according to direct labour bases, which made sense at the beginning of twentieth century but not in a new business environment characterised by intensive investment in advanced manufacturing technologies. As labour costs are now a small proportion of production costs then allocating costs on labour bases often brings incorrect allocation practices and erroneous product costing. Furthermore, they argued that developments in computer systems made calculations using more sophisticated allocation bases than labour ones more practical. They attributed the decline of MA to the development of accounting standards in the UK and USA which led MA to become financially oriented and geared to valuing stocks in financial reports for external users rather than providing managers with relevant data for decision-making. Johnson and Kaplan alleged that this subservience was perpetuated in management accounting teaching, particularly in business schools, that prioritised financial reporting. A similar conclusion was reached by Friedman and Lyne (1995 following an examination of post 1945 examination syllabuses of the Chartered Institute of Management Accountants in the UK). Nevertheless, this argument was largely polemic and lacked research backing. For example, field work in six large public UK companies found no evidence of MA subservience to financial accounting (Hopper et al., 1992).

Researchers do not disagree that conceptually ABC is an advance over traditional cost systems due to its assumption that 'activities consume resources and products consume activities' (Brimson, 1991; Brimson and Antos, 1994), and multiple cost drivers rather than traditional allocation bases drive costs (Cooper, 1990). In this respect, focussing on 'transactions' (Miller and Vollmann, 1985) brought greater realisation that the driver of overheads is the exchange of materials and/or information necessary to move production along - not necessarily the volume of production.

It is claimed that breaking down an organisation into activities creates possibilities for improving cost allocation, managing costs and even boosting overall efficiency (Raffish, 1991; Cooper and Kaplan, 1991, 92; Turney, 1991). Nevertheless, after the initial enthusiasm for ABC during the late 1980s and early 1990s, organisations and researchers have gained more experience of ABC and criticisms have emerged (Cobb et al., 1992; Innes and Mitchell, 1996; Noreen, 1991; Major, 2007; Major and Hopper, 2005; Hopper and Major, 2007). For example, though ABC allocates resources to activities and activities to cost objects through causal relations based on drivers, it does not guarantee that indirect costs are correctly attributed to products (Datar and Gupta, 1994) because some overheads incurred at a facility level are impossible to allocate (Cooper, 1990), and costs of activities can only be accurately allocated to cost objects when the relationship between them is strictly respected, i.e. costs associated with unit-level, batch-level and product-level activities should be allocated through unit-level, batch-level and product level bases respectively (Cooper, 1990; Noreen, 1991; Innes and Mitchell, 1996). In practice, due to organisational contingencies it is unlikely that such relationships will always be adhered to when choosing cost drivers (Innes and Mitchell, 1996; Cobb et al., 1992). Moreover, it is difficult if not impossible to achieve perfect homogeneity in activity cost pools (Innes et al., 1992; Datar and Gupta, 1994) due to problems of 'jointness' (which Kaplan later accepted as a problem that ABC could not solve). According to Noreen (1991), stringent conditions (including linear cost functions, zero fixed costs in cost pools, and no joint processes) must apply for ABC systems to provide relevant costs for product drop or product design decisions. It is tempting to surmise that ABC's advocates' switch from its potential for more accurate product costing to activity-based management relates to a recognition ABC's technical deficiencies for product costing (Hopper, 1994; Jones and Dugdale, 2002).

Furthermore, case studies have revealed difficulties in implementing and operating ABC (Malmi, 1997; Major and Hopper, 2005; Anderson, 1995; Hopper and Major, 2007; Friedman and Lyne, 1999). A decade ago the ABC literature was replete with methodological and operational concerns associated with economic measurement (Brimson, 1991; Brimson and Antos, 1994; Cooper, 1990), but now it is more concerned with issues associated with implementing ABC and its impact on organisations and individuals (Major and Hopper, 2005; Malmi, 1997; Argyris and Kaplan, 1994). ABC failure is often due to behavioural and organisational issues not properly addressed during implementation (Hankinson and Lloyd, 1993; Major and Hopper, 2005; Malmi, 1997). Of course change is rarely easy in organisations, especially when people are asked to change their beliefs (see Burns and Scapens, 2000) and ABC often requires a change of mind-set which often makes people reluctant to accept it (Major and Hopper, 2005; Major, 2007; Argyris and Kaplan, 1994). Argyris and Kaplan (1994) suggest that resistance to ABC can be alleviated by cultivating commitment through enabling individuals to personally deduce why they should commit energy and action to such a project. However, whilst not wishing to deny the usefulness of good change management practices, the problems go deeper than this. For example, Innes and Mitchell (1991) found managers often refused to adopt ABC since its implementation is costly in terms of human and physical resources and it involves considerable disruption. There is now evidence from empirical studies that ABC can generate dysfunctional consequences within organisations, perhaps because as Hopper argued (1994: 487) "systems perceived by managers to be potentially threatening can be rendered unworkable through managerial biasing and manipulation of data". This may explain why so few companies are adopting ABC given the widespread interest it engendered in management consultancy and academia. For instance, a survey of the UK's largest companies by Innes and Mitchell (1995) found that less than 20% had adopted ABC. Five years later, a repeat survey by Innes et al. (2000) concluded that the use of and interest in ABC had not increased and slightly more respondents had rejected ABC after assessment.

In sum, ABC is not a panacea that solves all costing problems. Kaplan and Cooper's initial case studies in early 1990s were important for popularising ABC amongst practitioners and in academia but they were published outside the normal scrutiny of academic refereeing. Some researchers suggest that ABC is little more than a fashion or fad[4] (Innes et al., 2000). Certainly interest in ABC was initially shaped by rhetoric from the managerial fashion-setting community, namely business mass media, consultants and business schools, claiming it represented norms of rationality and progress (Abrahamson, 1991, 96) despite sceptics arguing it was only old business practices reinvented ("old wine in new bottles"). The latter claims may account for Johnson and Kaplan's (1991) historical account of MA development (subsequently repudiated by Johnson) claiming that engineers and managers of big US manufacturing companies in the early twentieth century used managerial techniques like ABC only to be frustrated by the subsequent dominance of financial accountants.

Further research is needed for researchers to conclude whether ABC is merely a management fashion or a superior alternative to traditional costing systems with tangible benefits to managers. However, given the complexities of implementing and operating ABC models in practice, its implementation costs, and doubts about its technical accuracy there is reason to suspect that ABC may fail to live up to its claims like previous heavily touted systems such as Zero-based Budgeting and Planned Programmed Budgeting Systems (Hopper, 1994).

THE CASE STUDY

Marconi

Marconi is a Portuguese telecommunications company established in 1925 to provide telegraphic radio communications. Until its integration into Portugal Telecom Group in 2002 it focused on long distance telecommunications services. For more than seven decades Marconi had a monopoly regime, was one of the most profitable Portuguese companies, and acquired a prestigious image for engineering expertise.

Until 1974[5] Marconi‘s main activity was telecommunications traffic between Portuguese ex-colonies in Africa. After the Portuguese revolution, Marconi's strategy switched to internationalisation through programmes of co-operation worldwide, supported by technological modernisation of satellite and submarine cables. Except for its first ten years of activity (1926-36), and despite lost traffic to and from Africa following decolonisation, Marconi was profitable. In the late 1980's Marconi entered new business areas in telecommunications services and also information systems; electronics; and financial services and property. Simultaneously, it enlarged its activities in telecommunications, in particular public telecommunications (local and international) outside Portugal, maritime mobile services, telecommunications engineering, telephone and business directories, TV broadcasting, corporate communications, value added services, and research and development. By the early 1990s Marconi was an important component of the Portuguese economy, and its organisation structure reflected its internationalisation and diversification strategy.

Following the reorganisation of Portuguese telecommunications in the 1990s, Marconi was integrated into the Portugal Telecom Group and focused just on long distance telecommunications: a contract between Portugal Telecom and Marconi in 1996 gave Marconi the sub-concession to operate telecommunications infrastructures for international services using submarine cables or satellite technology whereby Portugal Telecom transferred its telecommunications traffic business with Europe and North Africa to Marconi, and in return received Marconi's non-long distance business, including its financial investments abroad. Thus, Marconi became the single public provider of international telecommunications services in Portugal but lost its other businesses to its parent company. Marconi kept its legal independence until 2002 when it was fully integrated into Portugal Telecom Group.

The introduction of competition into Portuguese telecommunications and Marconi's integration into the Portugal Telecom Group pressurised Marconi to institute a programme of organisational change. External consultants were hired and several managerial projects were launched bearing upon its MA and information systems including: a revised strategic plan; a new career evaluation system; strategic control benchmarks; programmes to make staff aware of the competitiveness of Marconi's business environment; an Executive Information System to provide senior managers with operational and strategic information; SAP; and replacing Marconi's previous MA with ABC. Adopting ABC was the major project in terms of human and financial resources but also by the time and importance Marconi's top managers gave it. A significant factor in adopting ABC were directives from the EU and Portuguese regulators recommending that operators with dominant market positions or holding concessions (Marconi was both) should use it to justify cost-oriented prices and subsidies. The regulators advocated ABC on the advice of consultants who then instituted ABC in several European telecommunication operators. In 1997 Marconi, hired external consultants who continued to advise the regulators to help implement ABC throughout the company. In March 1998 the first results were obtained and discussed by managers. In 2000, when this case study commenced, ABC was fully implemented and provided data to Marconi's managers, its parent company and the Portuguese telecommunications regulator.

Marconi's Activity-Based Costing System

The ABC system implemented was relatively complex because it had to provide cost data to ICP and Marconi's parent for regulation and accountability, whilst giving managers sound economic information to support pricing and investment decisions, and outsourcing strategies.

Once the consultants had selected the personnel to implement ABC, they instituted training programmes where consultants and Marconi's managers from all departments could discuss the objectives of implementing ABC and its design. This was followed by the definition of activities, which required identifying relevant cost objects by analysing Marconi's business segments. It was concluded that international telecommunication services had five business segments: fixed telephone; telematics (which include telex, telegraphy, data communication, and store and forward fax); leased capacity (which embrace MID - Marconi's Direct Ethernet, broadcasting – videocommunications, TV and radio - restoring, and leased circuits); alliances with other operators (for example with Concert); and other segments (communications through satellite, Inmarsat-C, TCR – telemetry, engineering services in cables and international projects).

Table 1. Marconi's ABC Activities (source: Marconi's Dictionary of Activities)

Activities:
1) Main Activities
1.1) Activities Oriented to Customers
• Defining strategies in Telecommunications business
• Researching and analysing new business opportunities
• Elaborating and controlling marketing plan
• Researching markets and customers
• Developing products and services
• Commercialising products and services
• Billing
• Management of customers' debts
• Maintenance of customers' services
• Assuring the quality of services
1.2) Activities Oriented to Network
• Following telecommunications network technology trend
• Planning network telecommunications
• Managing telecommunications technology development
• Developing and implementing telecommunications network
• Managing the use of network resources
• Operating traffic
• Operating infrastructures
• Restoring telecommunications network
• Preventive maintenance
• Corrective maintenance
2) Supporting Activities
• Developing and managing human resources
• Managing internal communication and information
• Managing financial and physical resources
• Managing the image and the firm's external relations
• Legal support

When the consultants and management accountants had identified the main priorities of ABC, they calculated costs in five areas: systems (including submarine and terrestrial cables, and satellites that used systems such as Eurafrica; Tagide; Ariane; Columbus; and Americas); tracks (the path between operators to support communication - 'Marconi – British Telecom', and 'France Telecom – AT&T' were two); tracks by services and carriers (e.g. British Telecom-Fixed Phones); carrier (e.g. France Telecom; AT&T; Telkom South Africa; British Telecom); and Products/Services. Calculating costs for the first four areas was easy once the costs of products and services had been established, which required detailed information about costs of systems, tracks and carriers.

Marconi's ABC system comprised 115 activities, 71 of which were main activities and 44 supporting ones. The main activities consisted of activities oriented to customers (35) and the network (36). Table 1 briefly describes these activities[6].

Marconi adopted labour hours as first stage drivers, i.e. drivers of resources. Costs of each Marconi department were allocated to activities based upon the disclosure of employees' time by activity. Every three months all employees (except the board of directors) had to fill a sheet indicating their time spent on each activity. This procedure was called 'PMO' (an abbreviation of 'Ponto de mão-de-obra') and it was crucial for calculating the cost of activities in Marconi's ABC system. After allocating resource costs to activities, the cost of the latter was allocated to equipment, systems, tracks, carriers, and product/services using second–stage drivers (i.e. activity drivers). Costs of activities not related to these cost objects were classified as 'common costs'[7] in the income statement. Exhibit 1 shows Marconi's MAS architecture and Table 2 lists the second stage drivers, i.e. drivers of activities as selected by the consultants and management accountants.

Direct costs were identified for: operators - the cost of using telecommunication capacities of international operators and correspondents; leased capacities - the cost of supplying basic transport infrastructure to wholesale and retail services; rent – for the sub-concession from PT for operating international telecommunications traffic; transmission - costs of systems of submarine cables, satellite, radio, and network; and commutation - costs of managing the network. The costs of operators, leased capacities, transmission, and commutation were calculated from traffic records from engineering systems in operational centres, and allocations of resource costs to systems through PMO. Other direct costs came from allocations of activities to cost objects, following ABC implementation. Following this, direct costs represented roughly 75% of total costs, whereas previously they were 70%. This reduction was much lower than anticipated by

Exhibit 1. Marconi's ABC Conceptual System (source: Finance and Administration Department — Marconi)

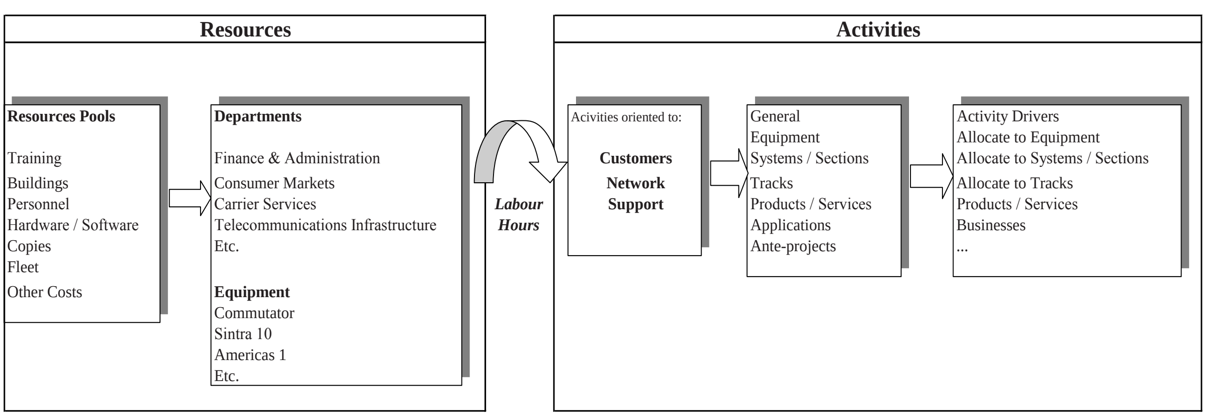


Table 2. Drivers of activities (source: Finance and Administration Department — Marconi)

	Description
Driver 1	Number of invoices valued by effort
Driver 2	Number of documents (incoming traffic) * 80% + number of open positions > 6 months * 20%
Driver 3	Number of invoices per product/service
Driver 4	Number of invoices * 70% + number of open positions > 6 months * 30%
Driver 5	Setups (Broadcasting – accidental services)
Driver 6	Number of processes valued by duration
Driver 7	Number of alterations in network valued by the type of service
Driver 8	Tracks valued by number of circuits
Driver 9	Similar allocation for the respective pseudo-department
Driver 10	Reallocation according to computer applications/machines DDS

consultants when they recommended ABC to the regulators and sometimes it entailed some heroic assumptions about cost causality.

Indirect costs consisted of joint and common costs. Joint costs of products (see footnote 6), carriers of traffic for a range of products/services, and capacity of transmission systems and commutation available but not used, were not significant being only approximately 2.5% of total costs. Common costs included costs not traceable directly to cost objects. They formed 22.5% of total costs and covered: activities not directly associated with cost objects (or a family of products); supporting activities not reallocable to primary activities; costs of capital of fixed assets not directly related to products/services; remuneration and fringe benefits of Marconi's Board of Directors, and personnel transferred to Portugal Telecom Group or new telecommunication firms; depreciation of fixed assets not directly associated with cost objects (e.g. equipment used by Marconi's Board of Directors, telecommunications systems not in use; and extraordinary costs.

Exhibit 2 presents the format of the income statement in Marconi's ABC system.

It details cost data for: fixed telephones (entering, in transit, and exiting to European countries, US, Canada, Central America, South America, South Africa, Angola, other PALOP's[8], other African countries, and Eastern countries); 'Marconi's Green Line'; telematics – telex, telegraphy, fax and data; leased capacity – national, broadcasting, leased circuits, restoring; and other products/services. Other income statements focusing on systems and activities were prepared in a format similar to exhibit 2, giving exhaustive details in categories 2.4.1. to 2.4.4., and 2.6.1. to 2.6.2. respectively.

The production department managed an important database[9] (SIGIR), which contained information crucial for inputs to the ABC system. In addition to inserting SIGIR, production department engineers were responsible for feeding other data into ABC including: cost of carriers, including data on capacity provided to Marconi and capacity used; identification of Marconi's operational centres' usage of commutation systems (namely, modem and multiplex systems); specific testing equipments' and air-conditionings' usage by transmission systems (i.e. cable, satellite and network) and commutation; Marconi's products/services usage of 'bubble' systems[10]; allocations by production department employees and managers of their time spent on activities; and allocation of resource costs (energy, depreciation, fuel, maintenance, reparation, etc.) to transmission and commutation systems. Without this information by the production department, Marconi's ABC

Exhibit 2. Income Statement in Marconi's ABC System (source: Finance and Administration Department – Marconi)

Description	Year (N)	Year (N-1)
1. Profits		
1.1. Traffic/Leased capacities		
1.2. Other profits		
Total 1.		
2. Direct Costs		
2.1. Costs of operator		
2.2. Leased Capacities		
2.3. Rent		
2.4. Cost of Transmission		
2.4.1. By Submarine Cables		
2.4.2. By Satellite		
2.4.3. By Terrestrial Cables		
2.4.4. Network		
2.5. Costs of Commutation		
2.6. Other Costs of Activities		
2.6.1. Activities Oriented to Customers		
2.6.2. Activities Oriented to Network		
Total 2.		
3. Joint Costs		
3.1. Joint Costs of Products		
3.2. Joint Costs of Carrier		
3.3. Capacity Available and Not Used		
3.3.1. Transmission		
3.3.1.1. By Submarine Cables		
3.3.1.2. By Satellite		
3.3.1.3. By Terrestrial Cables		
3.3.2. Network		
Total 3.		
4. Common Costs		
5. Total of Costs (2+3+4)		
6. Profit (1-5)		

system would not have been able to generate cost data.

The effective operation of ABC thus depended on co-operation between production engineers and management accountants. Consultants realised the importance of Marconi's production engineers to ABC's success and from commencement of its implementation tried to involve them in the project, though the engineers' participation was half-hearted, possibly because they resented how ABC symbolised their loss of power to the commercial departments[11] and the decline of production orientations following deregulation. Production engineers complained that ABC failed to serve their needs despite them being its main feeders. Consequently they frequently delayed submitting data, claiming they had insufficient time to meet accountants' demands for punctual information and to keep operational centres working efficiently simultaneously. They concentrated on the latter. This was exacerbated by job losses and work intensification. ABC also encountered resistance from workers, especially in the production department, who were reluctant to disclose information about activities, fearing it might curtail their autonomy and threaten job tenure. PMO and ABC became associated with increased workloads and meaninglessness. The dictionary of activities often provided a rough approximation of employees' work and did not capture important dimensions, so employees found the definitions and allocating labour time puzzling, which made completing PMOs difficult. Even if they were willing to accurately report the time spent on each activity, their recollections were hazy as they did this infrequently. Moreover, many production managers were overtly unenthusiastic about ABC from the outset. They never pressed workers to complete PMOs thoroughly or on time, their hostility to ABC was public, and they did not train subordinates to operate ABC. Hence, workers delayed allocating their labour time to activities and PMO returns were often inaccurate.

ABC data was included in the Executive Information System (EIS). When it was first introduced in 1998 it was intended to include only aggregated ABC data. However, because no system could systematically distribute detailed ABC data (except for special reports commissioned by managers), it was decided to provide such cost data to managers through EIS. However, as managers were reluctant to use EIS to obtain cost data, the management accountants resorted to preparing management and cost accounting reports as prior to adopting ABC. Their aim was to promote the use of ABC data by all departments and enhance enthusiasm for the new system but for more than two years management accounting reports were not prepared due to the management accountants' lack of time. Ultimately, after the complete liberalization of the telecommunication industry in 2002, Marconi's management accountants resumed preparing management accounting reports on an annual basis. They comprised: an executive summary of important events in Marconi and its business environment; income statements for the company, an analysis of costs/profits for the past and previous year; detailed analyses of each cost item in the income statement; and a comparison with the previous year. The analysis included: tables giving costs of activities by department; profits and costs of automatic telephone by traffic entering, in transit and exiting; costs of systems (satellites, and submarine cables located in Portugal or abroad); costs of commutation and networks; costs of tracks and carriers, amongst other elements; and conclusions. Appendixes were provided in electronic format.

After fully implementing ABC, Marconi began to use it to regularly provide the regulator and its managers with detailed costing data. Despite consultants' efforts to engage all Marconi's departments in ABC implementation, the managers most involved both in the implementation and usage of ABC were the two commercial departments of Marconi, which perceived it as valuable for it gave them relevant data for pricing and investment decisions. They were relatively unconcerned with the delays and possible inaccuracies believing ABC was superior to previous systems. In contrast, production personnel questioned the accuracy and slowness of ABC data and whether it helped secure cost reductions within operations. They saw ABC as clumsy, irrelevant, and a waste of their time. They were cost conscious but they used their own physical data for cost estimates. ABC might have met their needs had it been configured differently (Kaplan and Cooper (1998) claim that ABC can identify marginal costs accurately) but this was not attempted in the rush to meet the commercial departments' needs, which with hindsight was probably erroneous as ABC depended upon inputs from production.

Because of the problems facing ABC in Marconi, it is possible that this accounting system would not have continued in operation had the Portuguese regulator and the EU not imposed the periodical provision of cost accounting through ABC. However, since dominant telecommunication operators are obliged to send detailed cost data periodically to the national regulators and EU, it seems likely that ABC keeps in operation in the European telecommunication industry.

CONCLUSION

This chapter commenced by outlining debates on the efficacy of ABC. They suggested that under certain conditions ABC may be superior to other product costing systems but it is associated with a host of technical, implementation, and cost effectiveness issues. Ultimately ABC is another cost allocation system resting on arbitrary (but not necessarily unreasonable) assumptions about cost causation. Hence determining its ultimate merit may not reside in economic proofs but social consensus. Adopting ABC may have aided Marconi to improve its efficiency and competitiveness in newly competitive markets, though the reduction in the proportion of joint and common costs was marginal and failed to meet predictions. Nevertheless, managers from commercial departments were generally satisfied with the ABC system and used it for pricing or investment decisions, albeit with adjustments based on personal knowledge of Marconi's product costs and information on competitors' prices. On the other hand the production engineers were unimpressed and used ABC little, believing it was ill suited to operational decisions, was time consuming for little return, and was slow and inaccurate. Also the case study revealed a series of behavioural problems associated with implementing and operating ABC that had impacts on its accuracy. Thus there was empirical support from both sides of the ABC debate.

We concluded that ABC was implemented in Marconi to satisfy the needs of its managers and legal impositions from the Portuguese and EU telecommunications regulators. Adopting ABC helped Marconi to gain legitimacy from the market and its constituencies (in particular its competitors, customers, shareholders, and regulators) when setting interconnection prices to new operators and justifying subsidies for concessions because ABC was widely believed to be rational, accurate and reliable. This enhanced its attraction to regulators beset by external pressures over interconnection prices being fairly established and real competition occurring in Portuguese and European telecommunications markets. Furthermore, ABC enabled Marconi and its parent company to enhance the credibility of the cost accounting information provided to NYSE and exhibit to investors and others that they were efficient and 'modern' operators, which was crucial for gaining a renewed public concession contract with the Portuguese State and attracting new shareholders to the company.

REFERENCES

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review*, *16*(3), 586-612.

Abrahamson, E. (1996). Management fashion. *Academy of Management Review*, *21*(1), 254-285.

Anderson, S. W. (1995). A framework for assessing cost management system changes: The case of activity based costing implementation at General Motors, 1986-1993. *Journal of Management Accounting Research*, *7*, Fall, 1-51.

Argyris, C., & Kaplan, R. S. (1994). Implementing new knowledge: The case of Activity-Based Costing. *Accounting Horizons*, *8*(3), September, 83-105.

Brimson, J. A. (1991). *Activity accounting: An activity-based costing approach*. John Wiley & Sons, Inc.

Brimson, J. A., & Antos, J. (1994). *Activity-based management for service industries, governments entities and nonprofit organizations*. John Wiley & Sons, Inc.

Burns, J., & Scapens, R. W. (2000). Conceptualizing management accounting change: an institutional framework. *Management Accounting Research*, *11*, 3-25.

Cobb, I., Innes, J., & Mitchell, F. (1992). *Activity based costing: Problems in practice*. London: CIMA.

Cooper, R. (1990). Cost classification in unit-based and activity-based manufacturing cost systems. *Journal of Cost Management*, *4*(3), Fall, 4-14.

Cooper, R., & Kaplan, R. S. (1991). Profit priorities from activity-based costing. *Harvard Business Review*, May-June, 130-135.

Cooper, R., & Kaplan, R. S. (1992). Activity-based systems: measuring the costs of resource usage. *Accounting Horizons*, *6*(3), 1-13.

Datar, S., & Gupta, M. (1994). Aggregation, specification and measurement errors in product costing. *The Accounting Review*, *69*(4), 567-591.

Friedman, A. L., & Lyne, S. R. (1995). *Activity based techniques: The real life consequences.* London: CIMA.

Friedman, A. L., & Lyne, S. R. (1999). *Success and failure of activity-based techniques: A long term perspective*. London: CIMA;

Hankinson, H., & Lloyd, S. P. (1993). Change Management. In Brinker, B. J. (Ed.), *Handbook of cost management*. New York: Warren Gorham Lamont.

Hopper, T., Kirkham, L., Scapens, R., & Turley, S. (1992). Does financial accounting dominate management accounting – A research note. *Management Accounting Research*, *3*, 307-311.

Hopper, T. (1994). Activity based costing: A critical commentary. *Fukuoka University Review of Commercial Sciences*, *39*(1-2), 479-511.

Hopper, T., & Major, M. (2007). Extending institutional analysis through theoretical triangulation: Regulation and activity-based costing in Portuguese telecommunications. *The European Accounting Review, 16*(1), 59-97.

Innes, J., & Mitchell, F. (1991). *Activity based cost management: A case study of development and implementation*. London: CIMA.

Innes, J., & Mitchell, F. (1995). A survey of activity-based costing in the UK's largest companies. *Management Accounting Research, 6*, 137-153.

Innes, J., & Mitchell, F. (1996). A review of activity-based costing practice. In Drury, C. (Ed.) *Management accounting handbook*, 2nd edition. CIMA Publishing.

Innes, J., Mitchell, F., & Sinclair, D. (2000). Activity-based costing in the UK's largest companies: A comparison of 1994 and 1999 survey results. *Management Accounting Research, 11,* 349-362.

Johnson, H. T., & Kaplan, R. S. (1991). *Relevance lost: The rise and fall of management accounting*, 2nd edition. Boston: Harvard Business School Press.

Jones, T. C., & Dugdale, D. (2002). The ABC bandwagon and the juggernaut of modernity. *Accounting, Organizations and Society*, *27*, 121-163.

Kaplan, R. S. (1984). Yesterday's accounting undermines production. *Harvard Business Review*, July-August, 95-101.

Major, M. (2007). Activity-based costing and management: A critical perspective. In Hopper, T., Scapens, R. W., & Northcott, D (Eds.), *Issues in management accounting* (pp. 155-174), 3rd edition. London: Prentice Hall.

Major, M., & Hopper, T. (2005). Managers divided: Resistance and consent to ABC in a Portuguese telecommunications company. *Management Accounting Research*, *16*, 205-229.

Malmi, T. (1997). Towards explaining activity-based costing failure: Accounting and control in a decentralized organisation. *Management Accounting Research, 8*, 459-480.

Miller, J. G., & Vollmann, T. E. (1985). The hidden factory. *Harvard Business Review*, September-October, 142-150.

Noreen, E. (1991). Conditions under which activity-based cost systems provide relevant costs. *Journal of Management Accounting Research*, *3*, Fall, 159-168.

Raffish, N. (1991). How much does that product really cost? *Management Accounting* (US), March, 36-39.

Scapens, R. W. (1994). Never mind the gap: Towards an institutional perspective on management accounting practice. *Management Accounting Research, 5*, 301-321.

Shillinglaw, G. (1982). *Managerial cost accounting*. Richard D. Irwin Inc.

Staubus, G. J. (1971). Activity costing and input-output accounting. In Staubus, G. J. (1988), *Activity costing for decisions: Cost accounting* *in the decision usefulness framework*. Garland Publishing, Inc.

Turney, P. B. B. (1991). *Activity based costing: the performance breakthrough*. London: Kogan Page.

KEY TERMS

Activities: Consist of the aggregation of many related tasks (e.g. planning network telecommunications, operating telecommunications traffic, etc.).

Activity-Based Costing: A method that measures the cost and performance of process-related activities and cost objects through the assignment of costs to activities (first allocation stage) and activity costs to cost objects (second allocation stage).

Activity Cost Drivers: The allocation bases that are adopted to attribute activities costs to cost objects. They are also called second stage allocation bases (e.g. number of tracks valued by number of circuits, etc.).

Cost Object: Something for which a separate measurement of costs is desired (e.g. the cost of providing telecommunications services using satellites or submarine cables, etc.).

Direct Costs: Those costs that can be specifically and exclusively identified with a particular cost object (e.g. costs of transmission by terrestrial cables for fixed telecommunications services).

Indirect Costs: Those costs that cannot be specifically and exclusively identified with a given cost object (e.g. remuneration of the board of directors for calculating the costs of long distance telecommunications services, depreciation of the equipment used by central services, etc.).

Management Accounting: The provision of information required to formulate policies, planning and controlling the activities of an organisation, and to support decision-making processes and pricing strategies.

Resource Cost Drivers: The allocation bases that are used to allocate shared resources to individual activities. They are equally known as first stage allocation bases (e.g. labour time, metros square, etc.).

ENDNOTES

1 At that time the Portuguese telecommunications sector mainly was comprised of three firms established to provide services in specific fixed telephone services businesses - local/regional, national and intercontinental.

2 The Portuguese telecommunications regulator changed its name from ICP ('Instituto de Comunicações de Portugal') to ANACOM ('Autoridade Nacional de Comunicações') after the conclusion of this research.

3 The liberalisation of the telecommunications market brought opportunities for operators other than Marconi and its parent to use telecommunications infrastructures and fixed telephone networks that entered the 'public domain'. Until then they were used exclusively by Marconi and its parent. To use these infrastructures and networks, new operators must pay an 'interconnection rate' to Marconi (and its parent) based on costs determined from the latter's cost accounting systems.

4 Abrahamson (1996) describes these as "a relatively transitory collective belief, disseminated by management fashion setters, that a management technique leads rational management progress" (p. 257).

5 This was the year of the 'April Revolution', marking the end of the 13-year colonial war between Portugal and the African colonies, and the independence of those colonies.

6 These activities were sub-divided into more detailed activities, which are not included here for reasons of brevity.

7 In a rare number of cases there were additional costs of activities created by a family of products/services (e.g. fixed phones). These were included in the profit and loss account as 'joint costs of products', thereby being excluded from common costs.

8 This is an abbreviation for African countries whose official language is Portuguese.

9 SIGIR provides information about traffic and the network.

10 Specific commutation systems that are not automated and hence require identification of how they are used for products by engineers in the production department.

11 Marconi was comprised of two commercial divisions: The Consumer Markets department and the Carrier Services & Network Planning department. The Consumer Markets department was accountable for the development of new products and services, and for establishing the marketing strategies for the firm. The Carrier Services & Network Planning department was in-charge of planning the telecommunications network traffic and negotiating with carriers.

Chapter XIX

The Role of Organizational Culture to the Management of Telecommunication Companies: I. Background and Motivation

Antonios D. Kargas
University of Athens, Athens, Greece

Dimitris Varoutas[1]
University of Athens, Athens, Greece

ABSTRACT

This chapter explains what organizational culture is and analyzes its importance for the management of any company. Organizational culture must not be ignored during the decision-making process and managers must understand the existing culture of their organization in order to achieve their targets and to meet their goals. This chapter presents the theoretical link between organizational culture and a variety of variables, which affect organizational performance and efficiency, directly and indirectly. Such variables are knowledge management, organizational climate, leadership, quality, innovation and entrepreneurship, human resource management, and employee behavior. This chapter creates the starting point to study the link between culture and organizational strategy, enterprising practices and change management.

1.0 INTRODUCTION

During the last few years, a lot of different events, like the dominance of Japanese industry over the U.S. (Pascal & Athos, 1982), the fall of communism in Eastern Europe and China's economic reform (Dimitriades, 2005), have changed the shape of national economies and have formed a new global economy. At the same time, the technological, financial and political conditions have changed dramatically, as well as legislation and regulation in national and global level. All these have changed the traditional sources of competitive advantage (Jacome, Lisboa & Yasin, 2002), given rise to the risks undertaken, moved the core of research from the tangible to the intangible assets (Dimitriades, 2005) and demonstrated the role of **organizational culture** (Evans, Pucik and Barsourx, 2002).

Particularly, in order for every organization to succeed in its own national environment or abroad, it should be aware of its own culture (in order to recognize its strengths and weaknesses), the culture of the market (national or foreign) in which the organization wants to compete, and, of course, its competitors' cultures. Nowadays the managers not only need to know the existence of organizational culture, but they need to define its parameters which will allow them to deal with and manage culture, in order to succeed in areas such as organizational strategy and performance. So there is a growing interest towards the direction of finding out whether there is a relationship between management of organizational culture and the effective management of an organization. Despite the growing interest for organizational culture, as a means for achieving competitive advantage and superior performance (Evans, Pucik & Barsourx, 2002), there is still confusion about the way it affects organizations´ management.

This chapter will present in a comprehensive way, some of the most important aspects of the role of organizational culture to the management of companies in general. In Section 2, the most important definitions of organizational culture and its general importance for the understanding of any organization will be presented, while in Section 3 we will try to define the theoretical link between organizational culture and organizational variables, which are critical for the decision making process and the overall organizational management. Section 3 presents the indirect link between culture and organizational performance and efficiency, while in Section 4, the direct theoretical link between these elements will be presented. In the final section (Section 5) the conclusions of this chapter will be pointed out.

2.0 THE MEANING OF ORGANIZATIONAL CULTURE

Researchers with different academic and business backgrounds have come up with a lot of different theories about what organizational culture really is and how it has been formed over time. The variety of definitions is partly the result of the efforts of different researchers to explain organizational culture through their personal experience. So, they have used different data from different nations, societies, geographic locations and industries, aiming to explain organizational culture from their point of view. Also, many have been influenced by their personal and corporate history, the personalities of their employees, and their patterns of interaction.

The procedure of collecting data from different environments in order to describe and define a widely used term, like organizational culture, is acceptable. That is due to the fact that organizational culture is part of a whole system of cultures that affect the way the organization performs. It is part of what Straub *et al.* (2002) call "**virtual onion**".

This is not the only model trying to present the complex environment in which **organizational culture** is cultivated but is one of the most representative examples. The onion has seven layers and each one affects the employer through his everyday working process. The closer a layer is to the onion's core, the more it affects people and organizational culture is third in the row. Internal or external factors, such as the type of industry, the organizational structures, the national environment, are determining the influence that each layer has on the individuals of an organization. Virtual onion has been analytically presented by Karahanna *et al.* (2005).

The complexity of the issue makes it difficult to define precisely the organizational culture in a way that the produced definition could fit any kind of organization, any industry and any national economy. But, according to Rousseau (1990), all the different definitions describe the same thing from a different point of view, stating that it is not the definition of culture that varies so widely across organizational researchers, but the type of data researchers collect. It is, therefore, useful to know the most important definitions, because this will help us in understanding the meaning of organizational culture. The definitions which are presented are not the only ones that are considered to be of an importance for the understanding of the concept. They are some of the most known definitions and have affected a lot of researchers who have tried to connect organizational culture and effective management.

According to Schein (1985) "*corporate culture is the pattern of basic assumptions that a given group has invented, discovered or developed in learning to cope with its problems of external adaptation and internal integration and that have worked well enough to be considered valid and therefore to be taught to new members as the correct way to perceive, think and feel in relation to those problems*". This is one of the most often used definitions and it offers a dynamic view of organizational culture as well as a reason why it is a live part of every organization.

A new perspective for the meaning and the role of **organizational culture** has been given by Gordon (1991), who defines the concept as "*a set of shared meanings and understanding about the organization and its issues, objectives and practices*" (pp. 396 – 415). With his definition, Gordon introduced the use of organizational culture as a means for corporate control and behavioral management. Finally, Chu (2003) added an ethical perspective to the above definition by noting that "*organizational culture is the collective unconscious feeling of individuals in an organization about good or evil, normality of abnormality*" (pp. 505 – 509).

Because of the complexity and the difficulty to define precisely organizational culture, some researchers have compared it with the human character or personality (Park, Ribiere & Schulte, 2004). These researchers state that culture was the personality and the character of any organization, a variable that is very important and very wide to be defined. That was the reason why culture had no precise definition and was therefore ignored in the business environment (Hoffman & Klepper, 2000).

Even when it is not ignored, there are many business cases where organizational culture has been underestimated. That happens partly because organizational culture has a deep impact on the daily corporate operation. According to Mark (2006), if we would like to visualize culture, we could give it the shape of an iceberg. This **organizational iceberg** has surface and below surface components as well. The surface components are signs of the way that culture affects organizational reality, but they don't explain how this reality is shaped. The structuring device for this reality is below the surface, it is tacit and therefore not directly observable. This part of the iceberg is the largest and has a greater effect on the operation, the strategy, the managerial practices and the management of any kind of organization.

Even if it is impossible to find a widely acceptable definition for organizational culture, ignoring its presence may have a great consequence to the organization. Organizational culture affects directly and indirectly the operation and the management of the organization. In the next section, some of the empirical studies and the theoretical link between organizational culture and organizational variables and strategies will be presented in order to understand its great importance.

3.0 ORGANIZATIONAL CULTURE, VARIABLES AND STRATEGIES

Knowing that there is not a widely acceptable definition of organizational culture, many researchers have focused not on defining the concept, but on finding the positive effects of its effective management. That is because, it is well recognized that organizational culture has a significant role in the understanding of the way that different organizations are managed (Du Gay, 1996).

Of course there are also researchers who show distrust to the importance of organizational culture but this is something normal. Twenty years ago there were similar sentiments about concepts like quality, leadership, knowledge management and others. Progressively, these concepts have become necessary variables for the survival, the effective operation and the success of corporate goals. This change has shifted the centre of organizational analysis from financial results to non financial assets, like organizational culture.

Collins and Porras (1994, p. 55) suggest that "*profit is a necessary condition for existence and a means to more important ends, but it is not the end in itself for many of the visionary companies. Profit is like oxygen, food, water and blood; they are not the point of life, but without them there is no life*". Profit is truly important but developing procedures to achieve it for long - term periods is more important (Weymes, 2004). Pressure for short run profits may undermine the organizational capability to create long run and constant value (Kaplan & Norton, 1996) through elements like organizational culture (Banker *et al.*, 2004).

It is therefore necessary to explore the way and the extent to which organizational culture actively affects the everyday management. The influence of culture will be studied according to: a) a set of variables that affect the decision making process and the organizational performance and b) organizational performance and effectiveness.

3.1 Organizational Culture and Organizational Variables

There is a set of organizational non – financial variables, which have been proved to play a critical role in the decision making process and in the overall management of each organization. These variables are called by the general term "qualitative factors" (soft factors), because of the difficulty to be precisely defined and because their management is not an easy task as they are invisible organizational elements. The purpose of this section is to provide a theoretical link between these variables and organizational culture, in order to make clear the indirect effect of culture in the overall management of any kind of organization.

Such "qualitative factors" are:

a. The knowledge management and the organizational learning,
b. The organizational climate,
c. The effective leadership,
d. The innovation and the organizational entrepreneurship,
e. The effective management of human resources and
f. The employees' behaviour in sectors such as the trust and the commitment they show to the organization.

These variables are widely recognized by researchers around the world as critical for organizational success and will be analysed in the following paragraphs.

3.1.1 Knowledge Management and Organizational Learning

Knowledge and learning have been recognized as important variables for the organizational management (Sinkula, 1994) and as critical for the decision making process (Eisenhardt & Brown, 1998). Also, there is a positive link between knowledge and learning with the market competitiveness (Huber, 2001), the new product development (Madhavan & Grover, 1998), the innovation in general (Gloet & Terziovski, 2004) and the technological innovation in particular (Martin & Matlay, 2003).

Researchers of organizational learning and **knowledge management** confirm that these factors are strongly affected by **organizational culture** (Hutchings & Michailova, 2004). This influence is characterized by two counterbalancing forces. From one point of view, organizational culture can facilitate and strengthen the learning procedures and the knowledge management. From the other point of view, it can become an obstacle through the development of structures and operation which employees are unwilling to change.

The results of the researches from the first point of view are presenting organizational culture as a source of learning (Holsapple & Joshi, 2001) since they confirm that it supports the creation and distribution of knowledge. During the last two decades of 20th century, organisations that tried to create and manage knowledge, without cultivating an appropriate culture, have failed to reach their goals (Goh, 2003).

Researchers, such as De Long and Fahey (2000) have supported the idea of an organizational culture that encourages learning and the creation of knowledge. Garavan (1997) has argued that in order to incorporate learning in the daily organizational operation and processes, culture's synergy is needed through its values and motivation. Otherwise, it cannot be said that the organization is a learning one (Aksu & Ozdemir, 2005).

From another point of view, organizational culture can become an obstacle to the implementation of learning and knowledge management processes. McDermott & O'Dell (2000) have focused on the need to change organizational culture before the implementation of knowledge management programs. Other researches, in their attempt to find out the reasons of failure in managing learning processes, have concluded that organizational culture is the main obstacle to success (Tuggle & Shaw, 2000).

Dess & Picken (2000) highlighted organizational culture, through shared vision, as a necessary condition for the development of an organization that can learn, adapt, and respond effectively to a rapidly changing and competitive environment. The absence of shared vision has been analyzed as one of the most important causes of failure for the processes of organizational learning.

In 1997, Ernst & Young Center of Business Innovation conducted a study of 431 US and European organizations and identified culture as the current biggest impediment to knowledge transfer (Ruggles, 1998). The findings of this study were verified and were related to firm performance by the empirical results of other researchers (Ruggles 1998; DeLong & Fahey, 2000; Sveiby & Simons, 2002).

3.1.2 Organizational Climate

Organizational climate has been studied long before organizational culture, but its definition and the way it is supposed to affect organizational operation and processes highlight the meaning and the scope of organizational culture. According to Deshpande and Webster (1989), organizational climate describes what is happening in an organization, while culture describes why something is happening.

Nowadays, it is suggested that climate is a way that culture is expressed. Hale (2000) claimed that climate describes the organization in a particular moment while culture has a duration and can explain the organizational operation during a longer period. Currently, organizational culture is considered to be the greater framework in which appropriate climate is cultivated, as something more superficial (Flin *et al.*, 2000).

According to Schein (1983), **organizational culture** includes a variation of elements from different sections and from different mentalities, affecting and shaping organizational climate. When there is a dominant culture work environment, it tends to be more enjoyable and employees' morale boosts. This can lead to increased levels of sharing information, openness to new ideas and teamwork (Sadri & Lees, 2001). Also a well shaped organizational climate by a dominant culture, can become a powerful mechanism of attracting new executives from the market (Greger 1999), that can improve the organizational performance by transferring and diffusing their experiences.

Many researchers have used multi-trait matrixes in order to find the ways climate affects firm performance through the overall organizational culture. As part of organizational culture, climate has been found to affect the marketing strategy (Deshpande & Webster, 1989) and the overall strategic planning of the organization (White & Robicheaux, 1995). Organizational climate also affects the internal and external communication, as it affects the information flow and the context of interactions among employees (Falcione, Sussman & Herden, 1987).

3.1.3 Leadership

The relation between leadership and **organizational culture** has been heavily studied in the last twenty years, especially after the market globalization and the mobility of manager executives and employees. In the new globalized environment, the coexistence of different subcultures is considered to be a great challenge. The inability to understand and manipulate this diversity can lead to organizational failures in the sector of leadership as Helgstrand & Stuhlmacher (1999) have demonstrated.

A lot of researchers point out the existence of direct relationship between culture and the leadership style. These researchers are suggesting that values, ideologies and rules are shaping leadership style and that top managers tend to adapt these factors and to adjust their behavior (Schein, 1992). Of course, there are cases that particular leadership aspects transcend cultural boundaries and are universally accepted (House *et al.*, 2002). Schein (1990) recognized the link between culture and **leadership**. Leaders must respect the existing culture, but at the same time they should cultivate the appropriate and desired one. Through this procedure a new generation of managers can be developed from the internal organizational environment (Tichy and Cohen, 1997).

As Schein (1990) suggested at the initial stages of organizational life cycle, the leaders are shaping the culture. When organization matures, it is the culture that affects the managers and at later stages organizational culture can attract and develop the talents that are necessary to create managers. The direct link between organizational culture and **leadership** shapes a greater framework through which culture affects indirectly the overall organizational capacity to survive and to compete effectively.

Gestner (2002) supported the idea that successful organizations are characterized by powerful cultures, which are helping the effective management and the maintenance of organization in high levels of efficiency in a competitive environment. The role of leaders is to create the conditions for continuous changing of organizational culture by providing the example to all employees and by supporting the appropriate changing procedures. Even if leadership is an important organizational variable it must not been considered as a "magic elixir" for success, but it must be associated with organizational culture in order to achieve long term performance and growth (Pfeffer & Veiga, 1999).

It is recognized that there are limited empirical studies examining the relation between leadership and culture as well as their joint effect on important organizational outcomes (Xenikou & Simosi, 2006). One of the few empirical studies is the one Harris & Ogbonna (2001) conducted to investigate the relationship between leadership style and market orientation empirically. Their findings show a strong relationship between culture and leadership style.

3.1.4 Innovation and Entrepreneurship

In a fast changing environment, where globalized organizations play a critical role, the maintenance of any competitive advantage is harder than any time before. Organizations with greater interest in **innovation** and entrepreneurship will become more capable of maintaining their competitive advantages or developing new ones (Montes, Moreno & Fernandez, 2004).

During this process, the effective management of **organizational culture** can be of great importance to the cultivation of entrepreneurship and the development of innovative products and services. Especially in industries focusing on the development of new products, as the telecommunication industry, it is necessary to develop an innovative culture (Claver *et al.*, 1998). The importance of organizational culture relies on the strategic vision that encourages new enterprising ideas and innovative products (Pavitt, 1991). The misunderstanding of the importance of organizational culture may lead to failure in innovational attempts (Shane, Venkataraman & MacMillian, 1995).

Organizational culture permits managers to cultivate characteristics, quite different to the usual financial targets or performance measures. Thus, only through culture can we ensure that **innovation** procedures and entrepreneurship will have results for the organization (Hyland & Beckett, 2002). Otherwise, the fear of failure and the unwillingness to take technological risks will prevent the organization from maintaining any competitive advantage.

Numerous studies have produced evidence which highlights the importance of culture to organizational performance and efficiency, through the development of **innovation** and entrepreneurship. One of the most important studies is the one Deshpande *et al.* (1993) conducted. By using a synthesis of over 100 previous studies in organizational behaviour, anthropology and sociology, they defined four culture types and they showed that some of these cultural types are more likely to enhance innovation than others. Moreover, Ahmed (1998) used evidence from other researchers to conclude that companies aspiring towards innovative goals need to learn from the examples of highly successful companies like 3M, The Body Shop, Hewlett-Packard, Sony and Honda, whose leaders spend their energy and effort in building organizational cultures which perpetually create innovation.

3.1.5 Human Resources Management

The globalization of markets and the changes in the financial environment have affected dramatically the way human resources (HR) are managed (Schuler, 2000). The HR management has become a critical element for the success of every organization, whatever small or large, independently of the industry sector or national environment they belong to (Ulrich, 1997). The more effectively an organization manages its human resources, the more successful it is expected to be and the longer it will maintain its competitive advantage (Huselid, 1995).

The most fervent supporters of **organizational culture** believe that this idea is crucial for the management of human resources (Chan, Shaffer & Snape, 2004). The influence of organizational culture lies on the way it shapes the personal targets of managers and employees, on the way it determines the organizational tasks and on the way it engages human resources in the achievement of organizational goals.

The **human resources management** through culture is achieved because it connects top management with employees, by pushing forward a message according to which behaviors are expected to be in order to achieve organizational survival and growth (Maull, Brown & Cliffe, 2001). In this line, Deal & Kennedy (1982) suggested that employees, in organizations which have strong cultures, know exactly their tasks and duties, leading to increasing productivity. In contrast, when organizational culture is weak, employees waste time in order to decide how they must act, especially in cases of crises (Irani, Sharp & Kagioglou, 1997).

It can be concluded that organizational culture affects the qualitative management of human resources. It increases the cohesion among employees, it strengthens their ability to react in a desired way and it determines roles and tasks in the every day operation. The existence of a qualitative relationship between organizational culture and human resource management, creates difficulties in conducting empirical research, but it is considered a forthcoming field for research.

3.1.6 Behavior, Trust and Commitment

Organizational culture has been found to be related to employees' behavior and behavioral characteristics, such as trust and commitment to the organization. These elements are rather critical for the organizational success, as they encourage employees and they lead them to greater levels of participation to the organizational processes (Foy, 1994).

Organizational culture determines the acceptable and the expectable ways of behavior leading employees to the adoption of specified behavior models, in order to be incorporated in the organization. Kates & Robertson (2004) implied that every action in the organization is the result of its culture and they suggested that differentiations between culture and employee's behavior would have problematic situations as a consequence.

Considering that, it can be concluded that employees' behavior is part of the organizational culture, which affects their expectations and their productivity (Ivancevich, 2001). Moreover, culture affects the way employees conceive and face organization as an entity. Trust and commitment, as behavioral characteristics, are both describing how employees feel and react to the organizational structures and procedures.

Errol & Winston (2005) suggested that trust can be cultivated through the proper management of organizational culture, in order to avoid decrease in productivity that insecurity causes (Montes, Moreno & Fernandez, 2004). So, trust is part of organizational culture (Errol & Winston, 2005), and without the proper management there are few or no chances for the organization to achieve its goals and targets.

Respectively, Trice & Beyer (1984) related organizational culture with commitment, which removes the tendencies of resignation and absenteeism (Sheridan, 1992). Through the procedure of increasing the organizational commitment, the productivity increases (Holland, 1985), new employees and management executives are recruited (Judge & Cable, 1997) and the existing managers are willing to retain their jobs (Greger, 1999).

Generally, trust and commitment are depending on the acceptance and the degree the employees will adopt the organizational values and targets, that culture sets (Vakola & Nikolaou, 2005). Consequently, culture as a means of employees' socialization and successful incorporation to the organization, contributes positively to the cultivation of trust and commitment, so as to their satisfaction.

4.0 OVERALL PERFORMANCE AND EFFICIENCY

Organizational culture has been related to **performance** and efficiency from the beginning of 80s. This relationship was a result of an attempt to investigate the reasons why Japanese organizations had better outcomes than the American ones (Pascal & Athos, 1982; Deal &

Kennedy, 1982; Peters & Waterman, 1982). Also, there was a growing interest in the diversity of characteristics between successful and less successful enterprises operating in the same business environment (Lim, 1995).

Brooking (1996) suggested that organizational culture is a necessary element for the achievement of any organizational goal or target, while Itami (1987) argued that culture can be a useful mean for supreme performance and organizational efficiency. Quinn and Cameron (1983) relate culture with **performance** and efficiency through the life cycle of the organization. They argued that in different stages of the life cycle, there is a differentiation of the needs, the environmental conditions and prospects. So, in each stage there is a need for organizational culture with different characteristics in order for the desired outcomes to be achieved.

The qualitative character of the organizational culture has not prevented researchers from locating the negative consequences it can cause when it is ignored during the process of decision making. Kotter & Heskett (1992) described the relationship between culture and performance – efficiency in terms of profits or costs for the organization.

In 1998, "Fortune" magazine published a list of the leading enterprises for the year. At the top, there was General Electric, which has cultivating for several years a particular organizational culture, which offered top managers flexibility to decisions in order to make the organization capable of gaining any competitive advantage from market and industrial changes (Kahn, 1998).

Denison (1984) suggested that culture which empowers employees to participate in all the organizational procedures will determine the present and the future in financial performance. With his colleagues (Denison, 1990; Denison and Mishra, 1995; Denison *et al*., 2004), they have developed and empirically supported a theory of organizational culture and effectiveness that identifies four cultural traits that are positively related to organizational performance, namely involvement and participation, consistency and normative integration, adaptability, and mission. Theirs results were confirmed by Markoulides & Heck (1993), while some other researchers (Gordon & DiTomaso, 1992) argued that strong and dominant culture is related mainly to long term performance.

Walter (1985) presented organizations of high and low performance, from different industries, that have different cultural profiles, and Schein (1985) concluded that strong and dominant cultures lead to higher organizational performance than the weak ones. Newman & Nollen (1996) suggested that inability to manage organizational culture may lead to failures in terms of performance and efficiency.

All the above show the need to interconnect organizational values and behavioral rules (as they are shaped by culture) with targets and organizational strategy, so that desired performance and efficiency can be achieved (O' Reilly & Chatman, 1996). Moreover, other researchers (O'Regan & Ghobadian, 2004) consider organizational culture as a means for achieving supreme performance or efficiency, by helping the organization to exploit enterprising chances that arise from changes in the market or the industry, in which it operates.

This chapter aims to offer a clear picture of organizational culture as a powerful managerial tool, so that competitive advantage and supreme **performance** can be achieved (Barney, 1986). It is critical for every manager to know the existing culture and the desired one, in order to handle properly changes and dangers from competing in the global market.

5.0 SUMMARY

The knowledge of the organizational culture can be proved to be a critical factor in taking decisions about the future of any company. It is important for managers to know, not only, the present state of their organizational culture, but also the desired future condition, so that they can handle the indispensable changes. The purpose of this chapter was to show the importance of organizational culture for the effective management of any company.

Culture affects directly and indirectly the organizational performance and the management efficiency. The indirect effect is described by the link between organizational culture and: a) the knowledge management and the organizational learning, b) the organizational climate, c) the effective leadership, d) the quality in operation methodologies and in products and services, e) the innovation and the organizational entrepreneurship, f) the effective management of human resources and g) the employees' behaviour in sectors such as trust and commitment.

On the other hand, culture has also a direct effect on the organizational performance and efficiency. Its overall effect must not be ignored from top managers in order to take the proper decisions and to manage the organization effectively. Considering these, it is rather interesting to investigate the relationship between culture and organizational strategy, enterprising practices and change management. Moreover, it is critical to study how organizational culture affects telecommunication companies in their every day operation and in their management practices.

6.0 ACKNOWLEDGMENT

Authors would like to thank the two anonymous reviewers for their fruitful comments and suggestions towards the improvement of this chapter.

7.0 REFERENCES

Ahmed, P. K. (1998). Culture and climate for innovation, *European Journal of Innovation Management*, *1*(1), 30–43

Aksu, A., A., & Ozdemir, B. (2005). Individual learning and organization culture in learning organizations. *Managerial Auditing Journal*, *20*(4), 422-441.

Banker, R. D., Chang, H., Janakiraman, S. N., & Konstans, C., (2004). A balanced scorecard analysis of performance metrics. *European Journal of Operational Research*, *154*, 423–436.

Barney, J. B. (1986). Strategic factor markets. Expectations, luck and business strategy, *Management Science*, *32*, 1231-1241.

Brooking, A. (1996). *Intellectual Capital: Core Assets for the Third Millennium Enterprise*. London: Thompson Business Press.

Chan, L. L. M., Shaffer, M. A., & Snape, E. (2004). In search of sustained competitive advantage: the impact of organizational culture, competitive strategy, and human resource management practices on firm performance. *International Journal of Human Resource Management*, *15*(1), 17–35

Chu, K. F. (2003). An organizational culture and the empowerment for change in SMEs in the Hong Kong manufacturing industry. *Journal of Materials Processing Technology, 139*, 505–509

Claver, E., Llopis, J., Garcia, D., & Molina, H. (1998). Organizational culture for innovation and new technological behaviour. *Journal of High Technology Management Research*, *9*, 55–68.

Collins, J. C., & Porras, J. I. (1994). *Built to Last: Successful Habits of Visionary Companies*. London: HarperCollins.

Cook, R. A., & Lafferty, J. C. (1983). *Organizational Culture Inventory: OCI*. Plymouth, MI: Human Synergistics.

De Long, D. W., & Fahey, L. (2000). Diagnosing cultural barrieres to knowledge management. *Academy of Management Executive*, *14*(4), 113-127.

Deal, P., & Kennedy, A. (1982). *Corporate cultures*. Reading, MA: Addison-Wesley.

Denison, D. R. (1984). Bringing corporate culture to the bottom line. *Organizational Behavior, 13*, 5-22.

Denison, D. R. & Mishra, A. K. (1995). Toward a theory of organizational culture and effectiveness. *Organization Science*, *6*, 204-22.

Denison, D. R., Haaland, S., & Goelzer, P. (2004). Corporate culture and organizational effectiveness: Is Asia different from the rest of the world? *Organizational Dynamics*, *33*, 98-109.

Deshpande, R., & Webster, F. E. Jr., (1989). Organizational culture and marketing: Defining the research agenda. *Journal of Marketing, 53*, January, 3-15.

Deshpande, R., Farley, J. U., & Webster, F. E. (1993). Corporate culture, customer orientation and innovativeness in Japanese firms: a quadrad analysis. *Journal of Marketing*, *57*, 23-7.

Dess, G. G., & Picken, J. C. (2000). Changing roles: leadership in the 21st century, *Organizational Dynamics*, 28(3), 18-34.

Dimitriades, Z. S. (2005). Creating strategic capabilities: organizational learning and knowledge management in the new economy, *European Business Review*, *17*(4), 314-324.

Du Gay, P. (1996). *Consumption and Identity at Work*. London, Sage.

Eisenhardt, K. M., & Brown, S. L. (1998). Competing on the edge: Strategy as structured chaos. *Long Range Planning*, *31*(5), 786-9.

Errol, E. J., & Winston B. E. (2005). A correlation of servant leadership, leader trust, and organizational trust. *Leadership & Organization Development Journal*, *26*(1), 6-22.

Evans, P., Pucik, V., & Barsourx, J. L. (2002). *The Global Challenge: Frameworks for International Human Resource Management*. Boston, MA: McGraw-Hill Irwin.

Falcione, R., Sussman, L., & Herden, R. (1987). Communication climate in organizations. In Handbook of Organizational Communication: An Interdisciplinary Perspective F. Jablin et al., (Eds.), Newbury Park, CA: Sage Publications, Inc.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177–192.

Foy, N. (1994). *Empowering People at Work*. Ashgate, Aldershot.

Garavan, T. (1997). The learning organization: a review and evaluation, *The Learning Organization*, *4*(1), 18-29.

Gerstner, L. (2002). *Who says elephants can't dance*? New York, NY: Harper Collins. Inc.

Gloet, M.,& Terziovski, M. (2004). Exploring the relationship between knowledge management practices and innovation performance. *Journal of Manufacturing Technology Management*, *15*(5), 402–409.

Goh, S. C. (2003). *Improving organizational learning capability: Lessons from two.*

Gordon, G. G. (1991). Industry determinants of organizational culture. *The Academy of Management Review*, *16*(2), 396-415.

Gordon, G. G., & DiTomaso, N. (1992). Predicting corporate performance from organizational culture. *Journal of Management Studies*, *29*, 783-798.

Greger K. R. (1999). A positive corporate culture is the soul of retention. *Hotel and Motel Management*, *214*(17), 10.

Hale, A. (2000). Culture's confusions. *Safety Science*, *34*, 1–14.

Harris, L. C., & Ogbonna, E. (2001). Leadership style and market orientation: An empirical study.

European Journal of Management, 32(5/6), 744-764.

Helgstrand, K. K., & Stuhlmacher, A. F. (1999). National culture: An influence on leader evaluations? *The International Journal of Organizational Analysis, 7,* 153-68.

Hoffman, N., & Klepper, R. (2000). Assimilating new technologies. The role of organizational culture. *Information Systems Management, 17*(3), 36-42.

Holland, J. (1985). *Making Vocational Choices*, 2nd ed., Englewood Cliffs, NJ Prentice-Hall.

Holsapple, C. W., & Joshi, K. D. (2001). Organizational knowledge resources. *Decision Support Systems, 31*(1) 39-54.

House, R., Javidan M., Hanges, P., & Dorfman, P., (2002). Understanding cultures and implicit leadership theories across the globe. *Journal of World Business, 37*, 3-10.

Huber, G. B. (2001). Transfer of knowledge in knowledge management systems: Unexplored issues and suggested studies. *European Journal of Information Systems, 10*(2), 72-79.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38*, 635-72.

Hyland, P., & Beckett, R. (2002). Learning to compete: The value of internal benchmarking. *Benchmarking: An International Journal, 9*(3), 293-304.

Irani, Z., Sharp, J. M., & Kagioglou, M. (1997). Improving business performance through developing a corporate culture. *The TQM Magazine, 9*(3), 206–216.

Itami, H. (1987). *Mobilizing Invisible Assets.* Cambridge, MA: Harvard University Press.

Ivancevich, J. M. (2001). *International Human Resource Management*, 8th ed. New York, NY: McGraw-Hill/Irwin.

Jacome, R., Lisboa, J., & Yasin, M. (2002). Time-based differentiation – An old strategic hat or an effective strategic choice? An empirical investigation, *European Business Review, 14*(3), 184-93.

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology, 50*, 359-94.

Kahn, J. (1998). The World's most admired companies. *Fortune of 26 October,* 206.

Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Karahanna, E., Evaristo, R., & Srite, M. (2005). Levels of culture and individual behavior: An integrative perspective. *Journal of Global Information Management, 13*(2), 1–20.

Kates, S. M., & Robertson, J. (2004). Adapting action research to marketing. A dialogic argument between theory and practice. *European Journal of Marketing, 38*(¾),418-432.

Kotter, J. P., & Heskett, J. L. (1992). *Corporate Culture and Performance*. New York: The Free Press.

Lim B. (1995). Examining the organizational culture and organizational performance link. *Leadership & Organization Development Journal, 16*(5), 16-21.

Madhavan, R., & Grover, R. (1998). From embedded knowledge to embodied knowledge: New product development as knowledge management. *Journal of Marketing, 62*(4), 1-12.

Marcoulides, G. A., & Heck, R. H. (1993). Organizational culture and performance: proposing and testing a model. *Organization Science, 4*(2), 209-25.

Mark L. (2006). *Change, How to Do it and Make it Work, Tools and Techniques for Managing Change*. A Joint Working Publication.

Martin, L. M., & Matlay, H. (2003). Innovative use of the internet in established small firms: The impact of knowledge management and organizational learning in accessing new opportunities. *Qualitative Market Research: An International Journal, 6*(1), 18-26.

Maull, R., Brown, P., & Cliffe, R. (2001). Organizational culture and quality improvement. *International Journal of Operations & Production Management, 21* (3), 302-326.

Maxwell, G., McDougall, M., & Blair, S. (2000). Managing diversity in the hotel sector: the emergence of a service quality opportunity. *Managing Service Quality, 16*, 367-73.

McDermott & O'Dell (2000). Overcoming the cultural barriers to sharing knowledge. Available at: www.apqc.org/free/articles/km0200/index.htm

Montes, F. J. , Moreno, A. R., & Fernandez, L. M. (2004). Assessing the organizational climate and contractual relationship for perceptions of support for innovation. *International Journal of Manpower, 25*(2), 167-180.

Newman, K. L., & Nollen, S. D. (1996). Culture and congruence: The fit between management practices and national culture. *Journal of International Business, 27*, 753-79.

O'Regan, N., & Ghobadian A. (2004). Short- and long-term performance in manufacturing SMEs. Different targets, different drivers. *International Journal of Productivity and Performance Management, 53(5),* 405-424.

O'Reilly & Chatman, (1996). Culture as Social Control: Corporations, Cults, and Commitment, *Research in Organizational Behavior, 18*, pp. 157 – 200

Park H., Ribiere V. & Schulte W. D. Jr., (2004). Critical attributes of organizational culture that promote knowledge management technology implementation success. *Journal of Knowledge Management, 8*(3), 106-117

Pascal R. T., & Athos A. G. (1982). *The Art of Japanese Management*. Harmondsworth: Penguin Books.

Pavitt, K. (1991). Key characteristics of the large innovating firm, *British Journal of Management, 2*, 41-50.

Pfeffer, J., & Veiga, J. F. (1999). Putting people first for organizational success. *Academy of Management Executive, 13*(2), 37-48.

Quinn, R. E., & Cameron, K. (1983). Organizational life cycles and shifting criteria of effectiveness: some preliminary evidence. *Management Science, 29*,(1), 33-51.

Rousseau, D. (1990). Quantitative assessment of organizational culture: The case for multiple measures. In B. Schneider (Ed.), *Frontiers in Industrial and Organizational Psychology* (pp. 153-193). San Francisco, CA: Jossey-Bass.

Ruggles, R. (1998). The state of the notion: knowledge management in practices. *California Management Review*, 40(3), 80-9.

Sadri G., & Lees B., (2001). Developing corporate culture as a competitive advantage, *Journal of Management Development, 20,*, 10, pp. 853-859

Schein E. H. (1983). Organizational culture, *Organizational Dynamics 12*, 13– 28.

Schein, E. H. (1985). How culture forms, develops and changes. In Ralph, H.K., Mary, J.S. & Roy, S. (Eds), *Gaining Control of the Corporate Culture*. Jossey-Bass, San Francisco, CA.

Schein, E. (1990). Organizational culture. *American Psychologist, 45*, 109–119.

Schein, E. (1992). *Organizational culture and leadership*. A dynamic view (2nd Edition), San Francisco, Jossey – Bass

Schuler, R. (2000). The internationalization of human resource management, *Journal of International Management, 6*(3), pp. 239-60.

Shane, S., Venkataraman, S., & MacMillian, I. (1995). Cultural differences in innovation championing strategies, *Journal of Management, 21,* 931-52.

Sheridan, J. E. (1992). Organizational culture and employee retention, *Academy of Management Journal, 35,* 1036-56.

Sinkula, J.M. (1994). Market-information processing and organizational learning, *Journal of Marketing, 58,* 1, pp. 35-45.

Straub, D., Loch, K., Evaristo, R., Karahanna, E., & Srite, M. (2002). Toward a theory-based measurement of culture, *Journal of Global Information Management, 10* (1), 13–23.

Sveiby, K.E., Simons, R. (2002). "Collaborative climate and effectiveness of knowledge work-an empirical study", *Journal of Knowledge Management,* 6 No.5, pp.420-33.

Tichy, N. M. & Cohen, E. (1997). *The Leadership Engine: How Winning Companies Build Leaders at Every Level*, New York, NY, HarperCollins,.

Trice H. M. & Beyer J. M. (1984). *The cultures of work organizations*, Englewood Cliffs, NJ: Prentice - Hall

Tuggle F.D. & Shaw N.C. (2000). "*The effect of organizational culture on the implementation of knowledge management*", Orlando, FL, Florida Artificial Intelligence Research Symposium (FLAIRS).

Ulrich, D. (1997). Measuring human resources: an overview of practice and a prescription for results, *Human Resource Management, 36,* 3, pp. 303-20.

Vakola M. & Nikolaou I., (2005). Attitudes towards organizational change. What is the role of employees stress and commitment, *Employee Relations, 27, 2*, 2005, pp. 160-174

Xenikou A. & Simosi M., (2006). Organizational culture and transformational leadership as predictors of business unit performance, *Journal of Managerial Psychology* 21, 6, pp. 566-579

Walter, G.A. (1985). *Culture collisions in mergers and acquisitions,* in Frost, P.J. et al. (Eds), Organizational Culture, Sage, Beverly Hills, CA, pp. 301-14.

Weymes, E, (2004). A challenge to traditional management theory, *Foresight, Volume 6* Number 6, 2004, pp. 338-348.

White D. W., Robicheaux R. A., (1995). Linking Strategy Creation Styles With Corporate Climate In A Distribution Channel Context, available at: http://www.sbaer.uca.edu/research/sma/1995/pdf/01.pdf

KEY TERMS

Efficiency: The degree of how well a system or organization functions.

Entrepreneurship: The practice of starting a new organization, or starting a new enterprising activity in response to indentified opportunities.

Human Resource Management: The strategic management of employees so that the organizational objectives to be achieved.

Innovation: The introduction of new idea, method or device and its successful exploitation so that the organization increases its performance.

Knowledge Management: The way that an organization manages the information in order to create useful knowledge and that is distributed to the organizational members.

Leadership: The ability to affect human behavior so as to accomplish a mission designated by the leader.

Organizational Culture: The "personality" of an organization and is composed of the assumptions, values, norms, artifacts and behaviors of the members of the organization.

Organizational Learning: A theory about the way an organization learns through its everyday procedures.

ENDNOTE

1 Department of Informatics and Telecommunications, University of Athens, Athens, GREECE Tel: +30210-7275318, Fax: +30210-7275601, E-mail: arkas@di.uoa.gr

Chapter XX
The Role of Organizational Culture to the Management of Telecommunication Companies:
II. Applications and Case Studies

Antonios D. Kargas
University of Athens, Athens, Greece

Dimitris Varoutas[1]
University of Athens, Athens, Greece

ABSTRACT

This chapter analyzes the importance of organizational culture through a set of strategies and enterprising practices in order to emphasize the culture's necessity for the effective management of any organization and especially for the management of telecommunication companies. Moreover, the way culture affects the success and failure of a series of enterprising practices, such as mergers, acquisitions, strategic alliances and joint ventures is presented. All these create a rather interesting framework in which organizational managers should cultivate and implement the appropriate organizational culture in order to keep their organization competitive and well managed. Finally, case studies from the telecommunications industry will be presented, a field where culture plays a dominant role in either changing procedure or in developing new enterprising practices.

1.0 INTRODUCTION

In the previous chapter, the motivation for examining culture has been discussed. The direct and indirect relationship between culture and organizational performance and efficiency has been studied in order to understand its importance to the overall organizational management and to the decision making process

Performance is an important element for the organizational survival and growth but it is not the only one. The organizational efforts for improving financial performance can lead to the increase of short-term outputs but may also undermine the organizational capability for creation of future economic value (Kaplan & Norton, 1996). Only a healthy and balanced way of achievement of profits can help organization to maintain its competitive advantage and its financial performance for a long period.

Moreover, organizations are faced with great changes that have occurred in many areas of organizational operation and in many enterprising practices. Globalization of economies, liberalization of markets and the privatization of many state-owned companies has changed dramatically the way many organizations operate and compete with each other. New enterprise practices such as mergers, acquisitions and joint ventures have begun to be considered by the dominant companies or to take place already (Jackson, 2002).

In this environment a growing interest for the role of organizational culture has been established. Researchers have presented a link of culture with many aspects of organizational life other than performance. In an environment where risks are growing, well-prepared strategic plans can help every organization to avoid dangers to survive and compete. Moreover, the decision making process, which is related to the organizational strategy, affects the overall organizational performance and operation.

Secondly, organizations have adapted a new series of enterprising practices in order to compete successfully in the new globalized environment. These practices, particularly appropriate in the telecommunication industry, have led to new enterprising opportunities but also to new challenges. Implementing these practices may provide competitive advantage against any competitor.

After reviewing these organizational variables and practices and their relation to culture, the importance of culture for the overall management of any organization will be clear. However, there are only a few studies which investigated in which way organizational culture affects the telecommunication companies' management and their strategies. Considering this, it is interesting and challenging to study how organizational culture has affected telecommunication industry.

The rest of the chapter is structured as follows. In Section 2 the influence of culture on the formulation and adaptation of organizational strategy is studied, while in Section 3 its effect on a series of enterprise practices, applied particularly in the telecommunication industry such as organizational change, mergers and joint ventures, are analyzed. In Section 4 the role of organizational culture to telecommunication industry is discussed. Finally, in Section 5 the conclusions of this chapter and some trends for future research are presented.

2.0 ORGANIZATIONAL STRATEGY

Despite the disagreements about the definition and the context of organizational culture, many researchers suggest that its management is crucial for the decision making process and the formation of organizational strategy (Lee, 2004). In this section the culture's influence on strategy and decision making processes will be studied.

Pool (2000) and Daft (2001) consider culture to be a basic structural component of the organizational management system, while Valentine (2000) recognizes its importance for a set of organizational actions and for the organizational structure and the management philosophy. These researchers have been influenced by Deshpande & Webster (1989), who suggest that culture can be handled as any other organizational variable, in order to coordinate the decision making process and to diffuse the organizational strategy. Just like structure, strategy and technology, organizational culture can be managed so that higher levels of financial profits can be achieved.

Nowadays, most studies have focused on large organizations in order to clarify the link between culture and strategy (Barney, 1986). This research is following Hofstede's (1991) empirical results which demonstrated such a strong relationship between these two variables that it can be said that they are something inseparable. Therefore, culture can be cultivated as part of the **organizational strategy** in order to achieve the desire goals (Morgan, 1997) or else it may become a barrier to any change (Morgan, 1989).

According to Youngblood (2000) the most successful enterprises tend to treat their cultures as a critical factor for their success. Hankinson & Hankinson (1999) particularly studied the organizational cultures of the most successful enterprises and have found that top managers consider their organization as having a strong culture. It is thus interesting to investigate the exact way by which culture and strategy are related and how the decision making process is affected.

Firstly, organizational culture permits the differentiation of services. Organizations that have achieved this kind of differentiation tend to be more successful (Berry, 2000), as long as this differentiation is unique and provides extra value to the consumption of any product or service. A culture can make an organization unique not just because it makes the product or the service different, but because it differentiates the way it is offered to consumers.

If managers understand that culture is a mighty weapon against competitors, they can cultivate it as a new non – duplicated strategy in a series of issues such as marketing mix, quality in services and products, the price policy, the competitors' confrontation and others (de Chernatony & Cottam, 2006). Even though culture can provide competitive advantage, there seem to be organizations that ignore its presence. Even if organizational culture is considered an important aspect by top managers, it should be widely acknowledged by employees as well, in order to become a differentiative element and a competitive advantage (de Chernatony & Cottam, 2006).

Secondly, culture must fit to the organizational strategy, not only when strategy is shaped but also when it is applied (Hill, 1985). A culture is considered to be useful, only when it fits to the environmental and industrial conditions, in which the organization has to compete. The better it fits, the better the organizational performance is expected to be (Corbett & Rastrick, 2000).

On the other hand, even the best strategy may not be acceptable and therefore not exploited by employees if it does not fit the organizational culture (Irani, Beskese & Love, 2004). That is happening because culture shapes the framework in which strategy must be implemented and also because culture gives managers the means to empower the employees, in order to apply any desired strategic plan.

Of course the fit between culture and strategy is influenced by the dominance and the strength of culture in the organization. A strong and dominant culture is indirectly forcing managers to create strategic plans which fit culture, because they know by experience the employees' reactions to managerial plans. Generally, a strong culture in combination with a strong strategy can lead the organization to its goals and targets achievement. In contrast, a weak culture permits the adoption of any kind of strategic plan. Employees are indifferent towards the various strategies and they do not resist to the plans of top management. In this case the success depends solely on the strength and the vision that strategy can provide (Irani, Sharp & Kagioglou, 1997). The absence of a strong culture prevents the empowerment of employees, and strategy's strength may decline over the time.

It can be said that when culture and strategy fit, the forthcoming success strengthens even more the culture, which in turns increases its fit to strategy (Irani, Beskese & Love 2004). From all these it can be concluded that strategy and culture are inseparable and they affect each other. Only if managers consider them as a unit can the organization anticipate to gain the maximum from its strategic plans (Irani, Sharp & Kagioglou, 1997).

From all the above, it can be concluded that organizational culture has a great impact on the development and the implementation of organizational strategy. Moreover, it is essential for culture to fit strategy, so that their coordination allows the organization to gain maximum results. Finally, the bonuses from using culture as an alternative strategy must be considered in order for the organization to achieve and maintain a competitive advantage and supreme value.

3.0 ORGANIZATIONAL CHANGE AND NEW ENTERPRISING PRACTICES

During the last two decades of the 20th century the social, political and financial environment has changed worldwide. All these have increased the dangers that organizations should face. On the other hand, there are more opportunities for profits and competitive advantages. Organizations that are not focused on a permanent procedure of realizing the environmental changes might be put in danger. Any delay of recognizing risks and opportunities can prevent managers from proceeding to the necessary changes in order to reach a desired level of organizational operation. These changes include changes in organizational culture (Gilmore, Shea & Useem, 1997) and new enterprising practices, such as mergers and joint ventures (Isabella, 1990). All these aspects have been found to be positively related to culture through a procedure of shaping a new organizational identity (Gancel, Rodgers & Raynaud, 2002).

3.1 Organizational Change

Organizational change is considered to be an inevitable procedure that each organization will have to face for more than once in its lifetime. Successful enterprises are characterized by sensitivity to changes that occur in their operational environment (Chen *et al.,* 2004). So, it is important to understand the way in which organizations are managing changes in order to stay competitive and updated in relation to the developments in their industry sector (Johnson, 2004).

It is empirically proved that there are organizations that are not capable of responding to changes by shaping appropriately their operations, their structures and their management (Poole, 1998). The "*Economist*" magazine has focused on changing programs from all over the world and has found out that 85% of change efforts have failed. It has suggested that organization's change success depends on culture and the empowerment that can offer to employees. Of course, procedures, technologies and change models are useful tools in the changing trial, but without the appropriate culture all these cannot ensure successful changing management (Chu, 2003).

The first researches on the relationship between change and **organizational culture** have focused on their theoretical connection and on finding empirical evidence (Silvester & Anderson, 1999). At the same time, a dynamic research was developed, in order to make the bidirectional relationship between these two variables understandable (Silvester & Anderson, 1999). On the one hand, there is a growing interest about which culture is appropriate for the successful management of organizational change, while on the other hand there is interest about the way culture is reshaping during changes so as to make it feasible for the organization to adapt to new conditions.

Researchers have found that culture determines whether the organization can accept and encourage change (Irani, Sharp & Kagioglou, 1997). That is why the successful response to the international economic changes requires a flexible and sensitive culture (Elashmawi, 2000). Moreover, according to Karathanos (1998), managers should periodically analyze the relevance of values, which culture transmits, with organizational goals and flexibility to changes. In doing so, organization will be ready to face risks and to grab any given opportunity.

Problems may occur when culture does not have the appropriate characteristics to adapt to new conditions or to correspond to the required organizational changes (Balzarova et al., 2006). In these cases, the importance of organizational culture is becoming clear to managers due to the failures on implementing changing procedures (Clarke, 1994). That happens because culture is structured, collapsed and restructured on a daily basis and by itself, through the daily operation. This procedure is tacit and its result can not always be forecasted or controlled by managers (Knights & Willmott, 1995). This explains why a daily cultivation of the appropriate culture is needed instead of random efforts when the need for change appears as a necessity. In order for this to happen, managers should be aware of their existing organization culture and the desired condition, which the organization must reach.

This knowledge has multiple advantages for the organization. Firstly, it allows the localization of problems that may arise during the procedure of change from the existing to the desired situation. Secondly, culture elects the organizational strengths on which managers can rely on, not only during the change procedure but also during the future every day operations. Thirdly, it can forecast employees' resistance to change and to find ways to avoid this interception through a widely accepted manner.

All the above may determine the success or failure of organizational change and support the importance of culture in every procedure (Rashid, Sambasivan & Rahman, 2004). That is the reason why managers must know and understand the type and the characteristics of their organizational culture, before deciding about when and how they will implement change programs. In a different case there are increasing possibilities of failure during the management of organizational changes.

3.2 New Enterprising Practices

During the last twenty years, great changes have occurred in global economies. European Union became reality, communism has collapsed in Russia and Eastern Europe, and China's market has opened, while new technologies in telecommunication and transports made feasible the globalization of economies. In addition, political and legislative changes have developed the framework for deregulation and liberalization.

As a consequence of these changes, international markets extended their limits (Evans, Pucik & Barsourx, 2002) and a tendency of economic globalization has developed in Europe and in the rest of the world. This globalization has changed the traditional sources of competitive advantage (Jacome, Lisboa & Yasin, 2002) and developed a new enterprising framework in which organizations should adopt additional strategic approaches, in order to survive and to compete successfully (Dimitriades, 2005).

Such strategic approaches have been recognized to be the **mergers**, **acquisitions**, **joint ventures** and **strategic alliances**. In the case of strategic alliances and joint ventures, two or more enterprises contribute their financial, human, technological and physical resources or their expertise, in order to achieve competitive advantage. These enterprising practices emerged due to efforts for cost savings in operation, as was the case with GlobeOne, which has been the result of a joint venture between Deutsche Telekom, France Telekom, and Sprint (Deutsche Telekom, 2001).

Acquisitions are taking place when an organization is trying to buy out another one and to incorporate it progressively. Mergers are reflecting joint decisions through an effort for minimization of differences in operational and management structures. In the first case there is a dominant organization which imposes its management, while in the second case there is a creation of a new organization which has characteristics from both creators, after a progressive negotiation.

Nowadays, acquisitions and mergers are already considered mature enterprising practices, while joint ventures and strategic alliances are the appropriate tool to encounter the global market (Goldsmith, Walt & Doucet, 2000). Since 1985, these practices have begun to appear more often to already developed and later to developing economies. This increase is explained as a consequence of the way culture affects organizational structures and operation. When there are great differences between organizational cultures, acquisitions and mergers are more likely to fail than strategic alliances or joint ventures (Kogut & Singh, 1986). The globalization of economies in combination with differences in organizational cultures made these new enterprising practices a more common phenomenon (Hennart, 1988).

All the above are reflected in the high rate of strategic alliances and joint ventures, which have been measured to increase about 25% per year, after 1985 (Pekar & Allio, 1994), while a high rate of failure, around 50%, was recorded (Park & Russo, 1996). The reason for this high rate of failure must be addresed to the differences between the two cultures: the one that the organization, which enters a new market, has and the other the organization that already operates this market (Steensma et al., 2001). This problematic situation is happening because in the case of joint ventures and strategic alliances, one partner offers his knowledge about the market and the other one the structures and the management (Abdou & Kliche, 2004). All the above explain the growing interest for understanding the effects of organizational culture on acquisitions, mergers, joint ventures and strategic alliances.

Culture affects all the above enterprising practices in three ways in relation to the organizational operation. Firstly, culture's importance lies on its capability to create relationships and to develop networks (Hofstede, 1994). These networks include partners involved and theirs interactions, actions, distribution of resources, organizational climate and others. All these together are shaping the framework in which organization operates and they are affected by **organizational culture**. Each enterprising practice involves these networks and consequently managers should consider organizational culture as an important factor before taking any decision.

Secondly, it is generally accepted that market globalization and the creation of international companies have led organizations to a greater degree of demographic diversity (Williams & O' Reilly, 1998) and to greater heterogeneity of employees. Under these circumstances organizational culture can play a dominant role as an influential factor for the procedures and the relationships that employees develop (Williams & O' Reilly 1998). Employees are heavily affected by mergers, acquisitions, strategic alliances and joint ventures, as they have to get adapted to new operational conditions. That is why organizational culture must be managed properly in order to facilitate employees in their new tasks and responsibilities.

Thirdly, the globalization of economies has created the need for a new strategic orientation among organizations. International competition requires strategic vision not only to the organizational exterior contacts but also to its internal environment. This involves the development of customer orientation and focusing on trustful relationships between employees and external partners. Considering these, it is understandable why the entry to a new market, under any form of enterprising practice, requires the knowledge of existing cultures of the organizations which have to deal with and demands the development of a new culture which will lead to improved outcomes.

4.0 ORGANIZATIONAL CULTURE IN TELECOMMUNICATION COMPANIES

During the last 20 years, significant progress has been made in telecommunications internationally.

This progress has changed the scene worldwide. New providers of telecommunications services are coming along every day and most of them are coming from different business sectors. Thus, telecommunications have stopped being considered as fully regulated monopolies.

All these in addition to the technological changes and economic internalization have led to the increase of new risks and opportunities for organizations. This environment forced organizations to develop new products and services, to reorganize their structures and procedures, to cultivate innovative mentality, and to adopt new management models and organizational learning procedures. All these are evidence of the dynamic character of the telecommunication industry.

Moreover, telecommunication industry operates in a very competitive market and in such industries no organization can maintain its competitive advantage for long period without the effective management of organizational culture (D' Averi, 1994). The conditions that have been described above, plus the efforts for continuous competitive advantage, the necessity for flexibility at the decision making process, and failures in daily operation and in new enterprising practices have led to an increasing interest in the role of organizational culture in the telecommunication business.

Historically, during the "monopoly era", telecommunication companies have been based on operational expertise, paying attention only on issues such as finance, retail, wholesale prices etc. Under these conditions organizations have developed "silo–based" cultures which have removed them from goals such as performance, efficiency, customer orientation, human management procedures and competitiveness. Deregulation and liberalization have changed dramatically these conditions and many organizations have been forced to readjust their strategy, procedures and management. Moreover, telecommunication companies have realized the need for change their **organizational culture**.

In the next sections, case studies from the telecommunication industry, in which culture plays a determinant role will be presented. Efforts for cultural change along with cases of organizational failures as a consequence of ignoring its importance will be outlined. The aim is to give a picture of the research that has been done in this area and to reveal the need for further research and give some directions for future research on the relationship between organizational culture and effective management of telecommunication companies.

4.1 Case Study 1: Changing Organizational Culture

As discussed above, telecommunication companies have been forced to change their cultures as a result of deregulation of markets or privatization of companies. Moreover, many companies even if they were already operating in a competitive environment, realize that their culture is an obstacle in their efforts for success in competition and for better performance.

One example is Sri Lanka Telecom, a former stated – owned company which was privatized in 1996. In 1997, the Japanese NTT bought 35% of the company's shares and took over the management. As a consequence, structures have changed according to Japanese enterprising models (Sivananthiran & Venkata, 2004). Even if structures have been changed, culture has not, as it was clear by the behavioral patterns of the employees who have remained attached to past legacies of government bureaucratic administration. The employees did not share the top management vision and they had not internalized the organizational mission.

In order to operate effectively top management decided to change organizational culture through a procedure of social dialogue with employees' unions. This procedure has faced many problems from both sides but in the end there has been agreement on the steps that should be taken. The agreement led to a long–term procedure through which managers and union leaders came closer and were coordinated for the best interest of the organization. Managers and union leaders agreed that a customer oriented culture was necessary for the company. Employees were given training in communication and interpersonal skills, while a new value system was shaped and progressively declared. The new Chief Executive Manager, who was Japanese, helped the effort by removing communication restrictions between managers and staff, by introducing seminars and workshops, by encouraging creative ideas from employees and finally by introducing self-managed teams.

All these led to the cultivation of a new, customer oriented culture. The results of changing organizational culture, in terms of performance, are not known yet but their importance have been recognized from the telecommunication company's managers who used resources and time in a long – term and difficult procedure in order to change the bureaucratic culture of their employees.

This case shows the importance of organizational culture in order to enforce employees to share an organizational vision - mission and to achieve certain goals. Moreover, it is now clear that organizational culture can be managed by top managers in order to fit with their strategy, but only after long-term procedures. Another interesting conclusion is that after an acquisition, the new management must study the existing organizational culture and understand the employees' patterns before cultivating a new culture.

Another case is Telefonica Group which was a state-owned monopoly until 1998. By this time the organization had forty million clients, half of which were located in South America. Nowadays, Telefonica Group is consider as being one of the leading telecommunication companies worldwide and still maintains its leadership in Spain. Of course, the organization has faced serious problems after the market liberalization and its privatization. Competitors were pressing, the market shares were decreasing, profits were reduced, and clients were demanding better products and services.

The greatest challenge that Telefonica had to deal with was the transformation of its internal environment, in order to make clear that Telefonica was not anymore treating its clients as subscribers, but as customers who demand superior quality and lower priced products and services. This need for change in values and behavioral issues was reflected in the organizational statement for the new enterprising task, namely "*being identified by its current and potential clients as their favorite provider of telecommunications, voice, data, sound and image services*" (Claver *et al.*, 2000).

Telefonica faced these new challenges by trying to change its strong bureaucratic culture that has been established over seventy years of operation in a regulated environment. A regional Head Office was the starting point for changing culture, through certain management practices, which were implemented later in the whole firm. The efforts began with educational management courses given to the management board and to the top managers, so that they could have a clear picture of the desired customer oriented culture and of how they were expected to act. Moreover, the courses to the management board convinced everyone that the whole procedure was not a passing fashion.

Shortly, the courses were extended to middle line managers and employees, so that they could also understand how to do things right, from the customer's point of view, at the lowest possible cost. All employees and managers answered self-diagnosing tests which showed the progress that had been made and helped the commitment plan to be defined which was then applied to the whole firm. Nowadays, organization can claim that its goals have been reached, through the effective management of culture, in order to prepare employees for the forthcoming changes in market and the necessary changes to be done in behavioral terms.

These two cases give a clear picture about the role of organizational culture in many telecommunication companies. The need for **organizational change** has led them to the long-term and difficult procedures for changing their cultures, so that they can compete successfully in an unstable and competitive environment. The importance of managing their culture has been recognized through statements and proved by the resources that have been used in their efforts. Moreover, the important elements are that both organizations adopted the top-down change culture procedures which give a clue about the importance of visionary management. The employees and unions recognized and accepted the need for changing cultures in order to help organization to survive and to compete successfully.

However, there is a great difference between the two cases. Sri Lanka Telecom tried to change the organizational culture of the whole firm, while Telefonica first experimented on a regional Head Office, before taking any decision for the whole firm. In the first case, the firm changed management after an acquisition and the new management recognized the need for a customer oriented culture and was familiar with the ways this could be achieved. In the second case the cultural change was a decision of the existing management. The risk of failing led the managers to experiment on a regional head office. The two cases describe the need for changing organizational culture and how it was achieved by top management.

4.2 Case Study 2: Failing in Enterprising Practices

As it was discussed in Section 3, mergers, acquisitions, strategic alliances and joint ventures have been widely used as a form of strategic practice in order to improve organizational performance. However these managerial tools seem to have rather high rates of failure (Abdou & Kliche, 2004; Steensma *et al.* (2000). In order to avoid failure, further research on the role of organizational culture in relation to the creation and formation of the network has been recommended (Fletcher & Fang, 2006).

Telecommunication companies have applied these enterprising practices in many cases. Despite their best efforts, in many occasions they have failed to reach their goals and targets, and their strategic agreements have failed to ensure any competitive advantage. In these cases, organizational culture has been studied as a determinant factor, and it was stated that it organizational culture should be examined before any merger, acquisition, strategic alliance or joint venture takes place.

In 1999, two of the largest Nordic telecommunication companies attempted to merge, but their efforts failed. They were the Swedish telecommunication company, Telia and the Norway's largest telecom operator, Telenor. The two companies used to have elements in common as they were both former monopolies and state-owned, and operating in similar economies. There was also resemblance in their corporate practices. Moreover, they were operating in countries with similarities in language and national culture. All the above seemed to guarantee the merger's success, especially after the European Union's approval and the strong desire of both national governments.

Despite these facts, the **merger** was short-lived: only two months. The failure had cost approximately 200-250 million euro and both companies followed their own enterprising routes. Telenor remains Norway's largest telecommunication company, while Telia merged with the Finnish telecom operator Sonera in 2002 and became TeliaSonera, a leading telecommunication company in Nordic and Baltic (Fang, Fridh & Schultzberg, 2004).

The failure in merger has revealed the hidden differences in cultures. It has been recognized that the lack of preparation as a result of underestimation of potential differences and difficulties was one of the main reasons for this failure. Secondly, there seemed to exist lack of personal trust between the negotiating parties, which is partly a consequence of the inability to coordinate the differences in organizational culture. Thirdly, the "big brother versus small brother" syndrome has been detected – it made Norwegians feel ignored and underestimated, while the Swedish culture had not prepared them to deal with such situations.

This case study presents one of the most well known failures during a merging process of two telecommunication companies. The two parts had almost reached on agreement about the merge, as well as their national governments and European Union. Moreover, there was a perceived similarity, which led managers to underestimate actual cultural and subcultural differences (Fenwick, Edwards & Buckley, 2003). This was the point where negotiations failed. Today, it is recognized that organizational culture is a factor that should be examined before implementing any enterprising practice.

4.3 Case Study 3: Coordination Between Culture and Strategy

During the last 20 years, a significant progress has taken place in telecommunications internationally not only in terms of enhanced levels of technologies and telecommunication services but also in terms of changes in the structure of the traditional telecommunications market and the providers of telecommunication services.

In the EU the first steps towards deregulation of telecommunications and the formation of a common policy were taken in 1985. These were followed by the publication of the "Green Paper" in 1987, and a series of Guidelines and Decrees, which led to the implementation of this new common policy during 1987 – 1990. The deregulation of the telecommunication markets in the E.U. was the beginning of a new course for telecommunications, as new alternative operators have entered the liberated and deregulated markets. Especially, after 1998 all telecommunication companies in E.U. were facing international competition.

In this environment, Colt started operating in London metropolitan area in 1992. By 2003, it was operating metropolitan networks in 32 large cities, across 13 European states (Colt Telecom, 2002). Their interconnection has signaled the emergence of Colt as a pan – European operator. Colt's success was the result of its strategic behavior and its offensive culture.

Colt focused on higher technological quality and small costs by targeting on high-traffic customers who were in close proximity to each other (Turner, 2004). Small network and superior customer service were thus the key for strategic success, but was not the only factor. Colt had to compete with incumbents and new entrants as well, who could imitate its strategy or technology after some time. To gain a sustainable competitive advantage, Colt has cultivated an offensive culture, which relied on entering the market early and expanding rapidly before other entrants. At the same time, its staff should offer services of superior quality, so that the company could gain new customers and ensure long-term survival.

The coordination between culture and strategy led Colt to become on of the most successful providers in Europe for a long time. Financial problems as well as the increasing aggressiveness of incumbents against alternative operators, has lead Colt to change its culture and **organizational strategy** towards more defensive operating models. New culture was based on the need to secure position and market shares. After all, long-term customers were more profitable than temporary customers (Reichheld & Teal, 1996).

A variety of options emerged for Colt through the new defensive model. It could "sit and wait" (Turner, 2004), it could make strategic alliances or merge with a larger firm, or it could scale back its operations. Any of these options would mean changing the existing culture in order for the employees to understand the change in strategy. The new culture should focus not only on attracting new customers, but on expanding the relationship with the existing ones.

This case study presents how organizational culture is constantly changing as a consequence of the turbulent European telecommunication environment, which leads alternative operators to rearrange their strategy. It is unknown what would have happened if Colt had not coordinated its strategy and culture. On the other hand, it is clear that this coordination helped the firm to survive and become a pan– European operator.

5.0 CONCLUSION AND FUTURE WORK

The purpose of this chapter was to evaluate the importance of organizational culture and its contribution to the overall management of organizations and especially of telecommunication companies. It has been found that culture affects the development of organizational strategy, organizational change procedures, as well as a set of enterprising practices, such as mergers, acquisitions, joint ventures and strategic alliances. All the above develop a dynamic internal environment, which affect the decision making process.

Moreover, organizational culture has been found to affect many enterprising aspects of organizational life and operation of telecommunication companies. Three case studies have been analyzed, which present its importance for change management and for the successful enterprising practices. In all these cases, the necessity for leveraging on the existing culture in order to manage efficiently the organization has been acknowledged. However, there are not yet any widely accepted quantitative measurements that could provide comparable results.

As part of future work, it would be interesting and fruitful to study national telecommunication industry as a whole to understand the way that culture affects management and to compare different cultures that lead to different organizational performance and efficiency.

6.0 ACKNOWLEDGMENT

The Authors would like to thank the anonymous reviewers and the editor for their fruitful comments and suggestions towards the improvement of this chapter.

7.0 REFERENCES

Abdou, K., & Kliche, S. (2004). The strategic alliances between the American and German companies. A cultural perspective. *European Business Review, 16*(1), 8-27.

Balzarova, M. A., Castka P., Bamber C. J., & Sharp J. M. (2006). How organisational culture impacts on the implementation of ISO 14001:1996 – a UK multiple-case view. *Journal of Manufacturing Technology Management, 17*(1), 89-103.

Barney, J. B. (1986). Strategic factor markets. expectations, luck and business strategy, *Management Science, 32*, 1231-1241.

Berry, L. L. (2000). Cultivating service brand equity. *Journal of the Academy of Marketing Science, 28*(1), 128-37.

Chen C., Yu, C., Yang, S., & Chang, H. (2004). A customer-oriented service-enhancement system for the public sector. *Managing Service Quality, 14*(5), 414-425.

Chu, K. F. (2003). An organizational culture and the empowerment for change in SMEs in the Hong Kong manufacturing industry. *Journal of Materials Processing Technology 139*, 505–509.

Clarke, L. (1994). *The Essence of Managing Change*, Englewood Cliffs, NJ: Prentice-Hall.

Claver, E., Gasco, J. L., Llopis, J., & Lopez, E. A., (2000). Analysis of a cultural change in a Spanish telecommunications firm, *Business Process Management, Journal, 6*(4).

Colt Telecom (2002). *Annual Report.* Available at: www.colt.com.

Corbett, L. M., & Rastrick, K. N. (2000). Quality performance and organizational culture: A New Zealand study. *International Journal of Quality & Reliability Management, 17*(1), 14–26.

D'Aveni, R. A. (1994). *Hypercompetition: Managing the Dynamics of Strategic Maneuvering.* New York, NY: The Free Press.

Daft, R. L. (2001). *Organisation Theory and Design*. Cincinatti, OH, South-Western College Publishing.

de Chernatony, L., & Cottam S. (2006). Why are all financial services brands not great? *Journal of Product & Brand Management, 15/2* 88–97.

Deshpande, R., & Webster, F. E. Jr., (1989). Organizational culture and marketing: Definin the research agenda. *Journal of Marketing, 53*, January, 3-15

Deutsche Telekom (2001). Die Deutsche Telekom, France Telecom und Sprint sind von ihrem internationalen Joint Ventures Global One u¨berzeugt. Available at: www.telekom.de/dtag/presse/artikel/0,1018,x125,00.html (accessed 25 March).

Dimitriades, Z. S., (2005). Creating strategic capabilities: organizational learning and knowledge management in the new economy. *European Business Review, 17*(4), 314-324

Elashmawi, F. (2000). Creating a winning corporate culture: Experience inside the Asian telecommunications industry. *European Business Review, 12*(3), 148–156.

Evans, P., Pucik, V., & Barsourx, J.L. (2002). *The Global Challenge: Frameworks for International Human Resource Management.* Boston, MA: McGraw-Hill Irwin.

Fang, T., Fridh, C., & Schultzberg, S., (2004). Why did the Telia–Telenor merger fail? *International Business Revie 13*, 573–594.

Fenwick, M., Edwards, R., & Buckley, P. J. (2003). Is cultural similarity misleading? The experience of Australian manufacturers in Britain. *International Business Review, 12*(3), 297-309.

Fletcher, R., & Fang, T. (2006). Assessing the impact of culture on relationship creation and network formation in emerging Asian markets. *European Journal of Marketing, 40*(3/4), 430-446

Gancel, C., Rodgers, I., & Raynaud, M. (2002). *Successful Mergers, Acquisitions and Strategic Alliances: How to Bridge Corporate Cultures.* New York, NY: McGraw-Hill.

Gilmore, T. N., Shea, G. P., & Useem, M. (1997). Side-effects of corporate cultural transformations. *Journal of Applied Behavioral Science, 33*(2), 174-89.

Goldsmith, M., Walt, C., & Doucet, K. (2000). New competencies for tomorrow's global leader, *CMA Management, 202473*(10), 20-4.

Hankinson, P., & Hankinson, G. (1999). Managing successful brands: An empirical study which compares the corporate cultures of companies managing the world's top 100 brands with those managing outsider brands. *Journal of Marketing Management, 15*, 135-55.

Hennart, J. F. (1988). A transaction cost theory of equity joint ventures. *Strategic Management Journal, 9*(4), 361-74.

Hill, T. (1985). *Manufacturing Strategy,* Basingstoke Macmillan.

Hofstede, G. (1991). *Cultures and Organisations: Software of the Mind.* New York, NY,McGraw-Hill.

Hofstede, G., (1994). The business of international business is culture, *International Business Review, 3*(1), 1-14.

Irani, Z., Beskese, A., & Love, P. E. (2004). Total quality management and corporate culture: constructs of organisational excellence. *Technovation, 24*, 643–650.

Irani, Z., Sharp, J. M., & Kagioglou, M. (1997). Improving business performance through developing a corporate culture. *The TQM Magazine, 9*(3), 206–216.

Isabella, L. A. (1990). Evolving interpretations as a change unfolds: how managers construe key organizational events. *Academy of Management Journal*, *33*(1), 7-41.

Jackson, T. (2002). *Cross-cultural Approach*. London: Sage Publications.

Jacome, R., Lisboa, J., & Yasin, M. (2002). Time-based differentiation – an old strategic hat or an effective strategic choice? An empirical investigation. *European Business Review, 14*(3), 184-93.

Johnson, D. M. (2004). Adaptation of organizational change models to the implementation of quality standard requirements, *International Journal of Quality & Reliability Management, 21*(2), 154-174.

Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Karathanos, P. (1998). Crafting corporate meaning (developing corporate culture). *Management Decision, 36*(2), 123–132.

Knights, D., & Willmott, H. (1995). Culture and control in a life insurance company. *Studies in Cultures, Organizations and Societies, 1,*(1), 29-46.

Kogut, B., & Singh, H. (1986). Entering the United States by acquisition or joint venture, country patterns and cultural characteristics, Working Paper, Reginald H. Jones Center, Wharton School, University of Pennsylvania, Philadelphia, PA.

Lee, Y. (2004). Customer service and organizational learning in the context of strategic marketing. *Marketing Intelligence & Planning, 22*(6), 652–662

Morgan, G. (1989). *Creative Organization Theory: A Resource Handbook*. London: Sage.

Morgan, G. (1997). *Images of Organisation*. London: Sage Publications.

Park, S. H., & Russo, M. V. (1996). When companies eclipse cooperation: an event history analysis of joint venture failure. *Management Science*, *42*, 875-90.

Pekar, P., & Allio, R. (1994). Making alliances work. Guidelines for success. *Long Range Planning, 27*(4), 54-65.

Pool, S. W. (2000). The learning organization: motivating employees by integrating TQM philosophy in a supportive organizational culture. *Leadership & Organisation Development Journal, 21(*8), 373-8.

Poole, P. P. (1998). Words and deeds of organizational change. *Journal of Managerial Issues, 10*(1), 45-59.

Rashid, Z. A., Sambasivan, M., & Rahman, A. A. (2004). The influence of organizational culture on attitudes toward organizational change. *The Leadership & Organization Development Journal*, *25*(2), 161-179.

Reichheld, F. F., & Teal, T. (1996). *The Loyalty Effect: The Hidden Force behind Growth, Profits and Lasting Value*. Boston, MA: Harvard Business School Press.

Schuler, R., & Jackson, S. (2001). HR issues and activities in mergers and acquisitions, *European Journal of Management*, *19*(3), 239-53.

Silvester, J., & Anderson, N. R. (1999). Organizational culture change. *Journal of Occupational & Organizational Psychology, 72*(1), 1-24.

Sivananthiran, A., & Venkata Ratnam, C. S. (2004). *Social dialogue at enterprise level:*

Successful experients. International Labour Organization, available at: www.ilo.org/public/english/dialogue/ifpdial/downloads/papers/social_dia_entlevel.pdf

Schuler, R., & Jackson, S. (2001). HR issues and activities in mergers and acquisitions. *European Journal of Management, 19*(3), 239-53.

Steensma, K. H., Marino, L., Weaver, M. K., & Dickson, P. H. (2000). The influence of national culture on the formation of technology alliances by entrepreneurial firms. *Academy of Management Journal, 43*(5), 951-73.

Turner, C. (2004). Changing strategic behaviour in European telecommunications: the experience of Colt. *Info, 6*(4), 252-260.

Valentine, S. (2000). International person-organization fit: The role of national culture. *International Journal of Management, 17*(3), 295-302.

Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: a review of 40 years of research. *Research in Organizational Behavior, 20*, 77-140.

Youngblood, M. D. (2000). Winning cultures for the new economy, *Strategy and Leadership, 28*(6), 4-9.

KEY TERMS

Acquisition: Takeover of a company by another in a friendly or hostile manner.

Joint Ventures: An entity formed between two or more parties to undertake economic activity together. The parties agree to create a new entity by both contributing equity and they then share in the revenues, expenses and control of the enterprise.

Mergers: A combination of two companies into one larger company, mainly through stock swap or cash payment.

Organizational Change: The way of changing procedures, structures or enterprising activities.

Organizational Culture: Is the personality of an organization and is composed of the assumptions, values, norms, artifacts and behaviors of the members of the organization.

Organizational Strategy: The vision of the organization, mainly targeting on a stronger market position.

Strategic Alliance: A formal relationship between two or more parties to pursue a set of agreed upon goals or to meet a critical business need while remaining independent organizations.

ENDNOTE

[1] Department of Informatics and Telecommunications, University of Athens, Athens, GREECE Tel: +30210-7275318, Fax: +30210-7275601, E-mail: arkas@di.uoa.gr

Chapter XXI
Dynamics of Mobile Service Adoption*

Hannu Verkasalo
Helsinki University of Technology, Finland

ABSTRACT

Many case examples in the mobile market have indicated that the success of mobile services (e.g. Internet browsing, email messaging or streaming video playback with mobile phones) is difficult to predict. Different factors serve either as drivers or bottlenecks in mobile service adoption. The present chapter has covered earlier research on mobile service adoption and utilized a newly developed handset-based mobile end-user research platform in obtaining data from 548 Finnish panelists in 2006. The main research goal is to understand the process of mobile service adoption by extracting new kinds of data straight from handsets. In addition to descriptive results, a path analysis model is developed that explains mobile service adoption contingent on a given set of explanatory variables. The chapter finds that user intentions have a strong impact on consequent adoption of the service. What is more, perceived hedonic benefits from the service are the strongest factor driving user intentions to use the service. The perceived technical capability to use the service and the role of the surrounding social network explain little why early-adopter kind of independent users intend to use services. Interestingly multimedia services are strongly driven by newer more capable handsets and mobile Internet browsing benefits significantly from block or flat-rate (instead of usage-based) pricing plans for transmitted data. The chapter develops several indices that measure time-varying characteristics of mobile services. Calculated indices for a set of mobile services in 2006 suggest that different mobile services are currently experiencing different phases in their life cycle.

1. INTRODUCTION

Mobile services have evolved quite a lot from mere communication oriented services (circuit-switched voice, text messaging, voice mailbox) to today's multimedia, content retrieval, browsing and other advanced services. The mobile Internet (see Funk 2004) is emerging and the IP-based service delivery is likely to hit the mobile mass market domain very soon. Overlay networks existing already in the Internet (Clark et al. 2006) may have spill-over effects to the mobile industry. The mobile Internet scenario contrasts sharply with the dominant, vertically-oriented way of doing mobile business (see e.g. Karlson et al. 2003, Verkasalo 2007a and Vesa 2005). The emergence of the mobile Internet is driven by the wide-scale adoption of smartphones (i.e. converged devices) along with improvements in both cellular (GSM and 3G) and alternative (e.g. WiFi) radio networks. In terms of data services the same service evolution trends have been seen in the "wired" Internet earlier that can be seen in the mobile domain today. For example, the movement from messaging data services to static content (Web) and further to multimedia streaming can already be seen in mobile service studies (Verkasalo 2007b).

Amidst the rapid evolution of the mobile industry many commercial service failures have taken place. It is difficult to pinpoint the reasons behind successes and failures. Typically not one but many issues affect the adoption of a particular mobile service. The reasons can be categorized into two main categories. First of all, a commercial/technical perspective includes issues that relate to marketing, positioning, developing, implementing, delivering and timing of the mobile service. These factors include e.g. demand forecasting, pricing, positioning of the service in the service provider's service portfolio, promotional activities, creation of end-user awareness, service quality management, and strategic push of the service in the value-chain (i.e. distribution management). These factors are called as t*echnological* or *business strategic* in Pedersen (2001). Second, the end-user perspective deals with end-user related factors driving or inhibiting service adoption. This perspective is called as *behavioral* in Pedersen (2001). Factors under this perspective include e.g. service usability, social pressure, network externalities, contextual environment, consumption choices, the user's motivation and technical capabilities. The first perspective deals more with the producer side of the market whereas the second perspective deals with the demand side of the market. Drivers and bottlenecks for service adoption might emerge in either domain.

Even though many potential factors explaining successes and failures of mobile services can be identified, it is often difficult to test hypotheses in practice. No suitable empirical research approaches have existed earlier to provide actual usage data to study the dynamics of mobile service adoption. Accurate data from end-users can be nowadays acquired with a handset-based mobile end-user research platform that was introduced in Verkasalo & Hämmäinen (2007). The new platform provides accurate usage statistics along with flexible tools to deploy questionnaire studies. The present chapter attempts to provide descriptive results on mobile service adoption with data from Finland 2006. In addition, a path analysis model is built explaining the main drivers and bottlenecks of mobile service adoption based on empirical usage data and questionnaire studies.

2 EARLIER RESEARCH

2.1. Theoretical Models Explaining Technology Adoption

The adoption research can generally be divided into four categories:

- Diffusion research (market focus)
- Adoption approach (individual user focus)
- Gratification research (needs of users focus)
- Domestication research (consequence of adoption focus)

The diffusion research focuses on the market-level phenomena, and studies the diffusion of technology in the whole market. Adoption research, on the other hand, considers individual users as a focal research object. Gratification research contributes by analyzing the different kinds of benefits users seek from new technologies, and domestication research analyzes the role of new technologies in integrating to the every day life of people. Although this research paper mainly applies the statistical models introduced in the adoption research, elements from other research approaches are also applied in building the framework introduced in chapter 3.3. Therefore all these approaches are discussed now in detail.

First, Rogers (1962) introduced the idea of "***diffusion*** of innovations", and approached the adoption process in diffusion terms. In his research the adoption process follows a bell curve, suggesting that different kinds of people adopt new technologies at different pace. Only the most technology enthusiastic people adopt new products/services at first, the mass market being more cautious and thus adopting slower. The late-adopters are the most technology averse people, typically purchasing new technology only when it is inevitable. Rogers' research still serves as the general background for many kinds of adoption research. The diffusion research can be used in studying the emergence of new services. In Rogers (1995) it is argued that adoption is initiated by a new technology, after which the social setting and communication channels boost the diffusion. Rogers' theory has laid ground for many other research frameworks, for example Christensen's (1997) theory of "disruptive technologies" that take over dominant technologies by having a disruptive diffusion path. Time is the core factor in Rogers' idea of technology diffusion, as adoption (penetration of service) follows different patterns at different points of time after its introduction.

The second approach is the ***adoption perspective***, in which an individual user standpoint is taken. Each individual makes her own decisions in considering whether to try the service or not. A wide domain of research stems from adoption of information systems science, utilizing theoretical models developed a couple of decades ago. Earlier models attempting to explain adoption of technologies (particularly information system technologies and this is why the approach is sometimes called as IS adoption science) include the *theory of reasoned action, theory of planned behavior* and *technology acceptance model*.

A theory of reasoned action (Fishbein & Ajzen 1975) is based on the individual's attitude towards the action and subjective norm of the action (expected behavior of others in response to the individual's action). Together these determine the behavioral intention to use the technology. Ajzen later expanded the model; in a theory of planned behavior (Ajzen 1985; 1991 a third concept exists, namely the perceived behavioral control. This reflects the difficulty of performing the action. These models communicate that technology adoption depends both on the individual's own perceived benefit of performing the action and the social norm driven by people around the individual. All in all, these frameworks suggest that usage patterns not only depend on the individual's own capabilities and interests, but also on the sociological environment and norms in the culture.

Davis (1989) follows similar logic in his framework - a technology acceptance model (TAM). He distinguishes two concepts. First, the perceived usefulness reflects the expected benefits from using a certain technology. Second, the perceived ease of use reflects pretty much the same thing as the perceived behavioral control in the theory of planned behavior, i.e. how difficult it is to use the technology. In predicting information technology adoption the TAM model developed by Davis is the most used framework, and by 2000 more than 400 journal articles had cited the two original TAM articles (Venkatesh & Davis 2000). Almost all the information technology adoption articles that will be discussed later in this chapter stem from the TAM model. Despite its popularity, further development of the model has taken place. For example, the original model as projected below suggested that perceived usefulness and perceived ease-of-use mediate all external factors (e.g. demographics), though this is not always the case (see e.g. Burton-Jones & Hubona 2005). The original model typically explains 40% of usage intentions and 30% of actual use (see e.g. Venkatesh & Davis 2000 and Meister & Compeau 2002). Applications of the framework might in the best case achieve better explanatory power.

Holistic adoption models (see e.g. Pedersen and Thorbjørnsen 2003) deployed earlier in the mobile context and stemming from the TAM model are close to the research approach applied in developing the theoretical framework for this research. Holistic behavioral models typically utilize statistical methods such as structural equation modeling (SEM), and they derive from theoretical models illustrated previously (particularly Davis' research). In Nysveen et al. (2005a) the adoption approach utilizing variations of the TAM model with SEM analysis is called as information systems research, as most studies utilizing structural equation models from the adoption perspective deal with the adoption of ICT services and systems. The research done by e.g. Pedersen and Thorbjørnsen (2003) is used as a basis for the model of this chapter. The next section will discuss the theoretical approaches to study mobile services in detail. Because of e.g. increasing context-specific nature and various ubiquitous characteristics (Heinonen and Pura 2006; Rask and Dholakia 2001) mobile services should be considered carefully when applying earlier information systems adoption models.

The other approaches to study the adoption of technology include a uses and gratifications research approach (see e.g. Leung and Wei 2000; Höflich and Rössler 2001) and domestication perspective (Haddon 2001; Ling 2001; Skog 2002). The former approach (***gratification research***) deals with the gratifications that users look for when using mobile services. These gratifications can be either utilitarian (business value or direct utility) or hedonic (entertainment oriented) (Flanagin and Metzger 2001). The latter approach (***domestication research***) has close linkage to sociology, anthropology and ethnology. Sometimes domestication research does not focus on individual adoption only, but extends to the adoption from the cultural or societal point of view (see e.g. Ling and Yttri 2002). Domestication research attempts to tackle the consequences of service usage and the integration of the technology/services in the customer's every day life (Pedersen 2005). These two frameworks reflect softer scientific disciplines in studying mobile services than the diffusion or adoption perspectives. All approaches share same concepts and ideas, and they are thus not totally separate from each other.

2.2. Emergence of Mobile Services

Little research on the emergence of mobile services exists. This section briefly describes some of the main outcomes of earlier research, particularly in light of this chapter. Many of the focal services studied in this chapter have just been introduced to the market, and no research on them is available. However, this section attempts to emphasize generic observations in the adoption of mobile services that have relevance also with new emerging services.

The mobile Internet is defined in this chapter to consist of new packet-switched mobile data services. The first mobile data services were hyped quite a lot in the public in the late 90s, but in practice the mobile Internet has not kicked off yet (Saarikoski 2006). Particularly the WAP technology and mobile email can be considered failures (Sigurdson 2001). Many reasons are suggested as possible bottlenecks, from pricing (see e.g. Gao et al. 2002) to general difficulties in terminal configuration (Verkasalo 2007b). These reasons were found as bottlenecks in mobile payment adoption, too (Mallat 2006b). MMS messaging experienced similar disappointing customer adoption rates at first to WAP, email and electronic payments are experiencing right now. In Japan NTT DoCoMo has achieved satisfactory demand for mobile Internet services, and Japan can be considered as the world's leading mobile market in many other dimensions, too (Saarikoski 2006; Minges 2005). The key difference to Western operators is that NTT DoCoMo has bundled service interfaces directly to the handset and actively pushed new value-added services. Consequently both the customer awareness has been increased and ease of adoption significantly improved.

The Internet service expansion to the mobile domain could serve as a prospective spark that could significantly push the extent of mobile data service usage. The reasons for the potential impact of Internet services can be divided into two. First of all, the pricing/commercial models of Internet services are different from operator-based mobile services (in terms of pricing, for example, Internet services are free of charge, or sold in flat-rate bundles). Second, the spill-over effects and value already embedded in Internet services (e.g. instant messaging communities, the variety of WWW-based content services, webmail) serve as a strong force if suitable mobile access technologies (e.g. the network access and adequate terminals for the usability's sake) can be deployed with low price. Earlier research has pinpointed the need for alternative radio access methods to fixed Internet service access, the key alternative being mobile cellular connectivity (Kearney 2001). These two factors (commercial/pricing models and mobile extension of the Internet) are evident in the final conclusions of Aarnio et al. (2002), as they state on the future prospects of mobile services that "*…prices must come down to overcome the critical mass threshold and in most cases the services should be integrated to the Internet...*".

Mallat (2006a) illustrates some interesting characteristics of mobile services in her dissertation. According to her studies, several successes and failures could be identified in the mobile payment service sector. Mallat concludes her dissertation with the note that usage situation and context are important in explaining the adoption of mobile electronic payments. The same can be generalized to the wider mobile domain. In studying mobile service adoption the special characteristics of mobile services – the freedom of context and location – should be internalized in the framework is possible. Mallat criticizes earlier mobile service adoption research for not taking contextual factors that well into account. Also Hejden et al. (2005) emphasized the importance of context in influencing the perceived value of services.

In tackling the drivers and bottlenecks of mobile services, Aarnio et al. (2002) studied 1 553 Finnish respondents and found five user segments from the sample that resembled classical adoption categories. Their paper studied both traditional Internet and new mobile services. The authors emphasize the important role of prices in mobile service adoption. Furthermore, they suggest that social norm should not be overlooked in mobile services. However, Nysveen et al. (2005a) suggest that different mobile services depend on explanatory factors differently. For example, some services are more dependent on social pressure/ norms than others. For example in the research of Pedersen and Thorbjørnsen (2003) the social norm did not explain that much of the variance.

Pedersen and Thorbjørnsen (2003) developed a structural equation model incorporating motivational, attitudinal, social and resource-related influences on the intention to use mobile services. Their model explained about 62-75% of the variance in the dependent variable (intention to use). In all of their case studies (focusing on different services) they found that both extrinsic (utilitarian) and derived (i.e. expressiveness) motivations have an important role to play, whereas intrinsic (entertainment-oriented) motivations are less explanatory. Pedersen (2005) continues by suggesting that many of the factors suggested in earlier ICT adoption research work fine in the context of mobile services, but also many new dimensions should be taken into account (such as expectations and subjective norm). Pedersen

(2005) argues that by extending the decomposed theory of planned behavior (Taylor & Todd 1995) with elements from domestication research the explanatory power of the models in mobile services can be increased. To criticize existing statistics models Nysveen et al. (2005b) argue that gender could have moderating effects in the adoption of mobile services, though the original TAM model suggests that all background variables should be fully mediated by the included explanatory factors. All in all, in statistically evaluating the determinants of usage, different services should be treated individually and one should be careful with cross-service generalizations.

Some mobile services are more goal-directed (utilitarian) instead of experimental (hedonic) (Nysveen et al. 2005a). Pedersen and Thorbjørnsen (2003) acknowledge the same. Also Wong and Hiew (2005) argue for the emphasis on the expected benefits from mobile services. Hejden et al. (2005) comply by concluding that utilitarian and hedonic values have some correlation but it is particularly utilitarian value that has a significant impact on the intention to use the services whereas hedonic values relate to more ad-hoc use cases. The utilitarian importance might have, however, derived from the market context at the time of the study in Hejden et al (2005), as many new mobile services emerging today are actually quite entertainment instead of business-oriented. Most new mobile services are therefore targeted at consumers instead of business customers. Hejden et al. (2005) nevertheless found that perceived risks from the service negatively drive the service's utilitarian value, whereas hedonic value is not affected by the perceived risks (e.g. the service is justified to be more difficult to use if it relates to "killing time"). Some papers (e.g. Tseng et al. 2007) emphasize that in addition to direct value also network externalities matter in analyzing the benefits of electronic services. In mobile instant messaging, for example, the value embedded in the buddy network affects the utility from using mobile instant messaging.

Pedersen (2005) notes that adoption studies in the mobile domain might have differing objects of research. He distinguishes between users (Green et al. 2001; Bakalis et al. 1997), services (see Verkasalo & Hämmäinen 2007; Kim 2001; Pedersen et al. 2001) and terminals (see Chuang et al. 2001; Skog 2000). In this chapter the approach is mainly from the user point of view, and the process through which end-users adopt services is the focal research objective.

3 EMPIRICAL ANALYSIS

3.1. Research method and dataset

A pioneering mobile end-user research platform was utilized in acquiring data for this research. The new mobile end-user research platform is based on a developed Symbian/S60 smartphone client that observes all kinds of usage actions taking place in the handset. Usage-levels stamps on any application, network or user interface level action is logged with accurate context-specific information (e.g. time). This data is sent to centralized servers every night for analysis purposes. Research is conducted in panels lasting typically 2-3 months, and a typical panel includes hundreds of interested customers which in the panel become panelists. All panelists participating in these study panels sign a contract and they are aware of the research process. Usage data is complemented with various WWW-based questionnaires through which to acquire data on issues that are not usage-related (e.g. motivations and attitudes). Though the accuracy and scope of acquired data is a clear contribution in the world of end-user research, the challenges include e.g. biased end-user domain (early-adopter users) and sample size (being still in the range of 400-1000 panelists). The research method has been used already in various papers, such as (Verkasalo 2006a, Verkasalo 2006b, Verkasalo & Hämmäinen 2006, Verkasalo 2007a, Verkasalo 2007b, Verkasalo 2007c and Verkasalo 2007d).

For more information on the research method, see Verkasalo & Hämmäinen (2007).

In acquiring data for this research SMS recruitment messages were sent to 28 000 Finnish consumer subscribers who owned a Nokia S60 device. 1 071 (3.8%) customers visited the recruitment site and answered the beginning questionnaire. Out of the registered panelists 695 (65%) managed to generate at least three active weeks of smartphone usage data. The rest either did not manage to install the research client or then quit the panel study earlier than was supposed. Some people might have used the smartphone with the research client installed as a secondary phone, and thus they were excluded, too. Out of the active panelists 548 (79%) answered the final questionnaire, which was more comprehensive than the beginning questionnaire. In addition to demographics, the beginning and final questionnaires covered various background questions related to user motivation, usage patterns and opinions on usability and performance of mobile services.

The details of the panel are introduced in Verkasalo (2007b). In brief, most panelists were young male consumer customers. The panel consisted of early-adopter customers, as in earlier handset-based service studies (see Verkasalo 2005). 37% of panelists considered themselves either experienced or very experienced smartphone users, whereas only 19% considered themselves as beginners. 56% considered themselves as normal in terms of smartphone usage patterns.

The questions asked from the panelists in background questionnaires are listed in Appendix A. The whole dataset used in this chapter consists of a questionnaire that was filled in by the panelists, together with a comprehensive set of aggregated usage-level measurements that reflect objectively panelists' actual behavior during the panel. The background questionnaire was used in obtaining data on the factors that have been found important earlier in studying technology adoption (see particularly Pedersen's research). Because of respondents' limited time, only one question was asked for each of the identified factors. Earlier statistical adoption studies have had more comprehensive questionnaires, and therefore they have the possibility to use both the measurement and structural parts of structural equation modeling. In this chapter only the path model is estimated. Data is collected for the services that were identified as being of great interest at that point of time (in 2006).

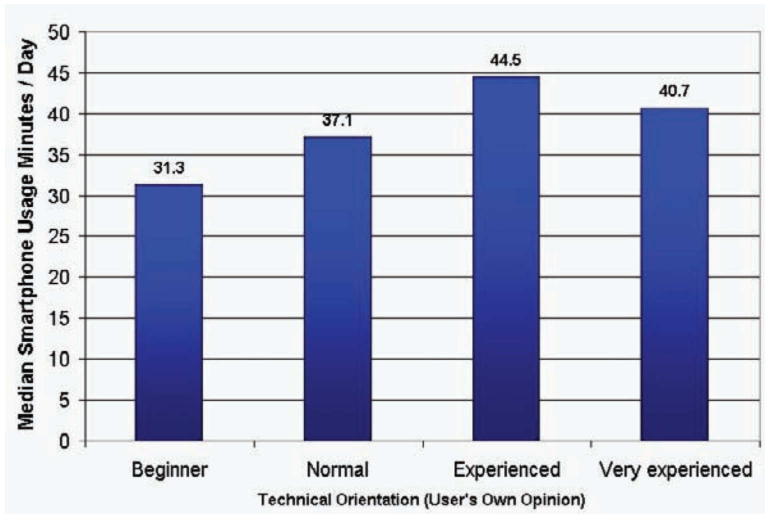
For each of these services a usage frequency variable was calculated measuring the share of days when a particular service was used out all active days spent in the panel. If a service is in constant use, this variable ranges from 50% (meaning usage every second day) to 100% (usage every day). If usage is less frequent, the usage frequency variable takes lower values. This variable well reflects the extent of use, and it is restricted to take values in the range of 0% to 100%. This variable is used as an indicator of usage in the estimated path model later in this chapter.

3.2. Descriptive Statistics

Figure 1 depicts the user's own opinion about himself/herself in terms of technical experience with smartphones and estimated average minutes of smartphone service usage per day. More experienced users spend time with smartphone devices more than less experienced users by observation. People with stronger technical interests and capabilities should be more likely to explore new services on the one hand and more likely to learn using them and adopting the service into every day use, on the other hand. Therefore the observed results on total smartphone usage time are quite expected. Because no data on user experience was available for all of the panelists, the total smartphone usage minutes per day is used a proxy for user experience in path modeling later in this chapter.

Next the focus is on particular services for which a comprehensive dataset was collected combining both questionnaires and actual usage

Figure 1. Correlation of technical orientation and extent of handset use

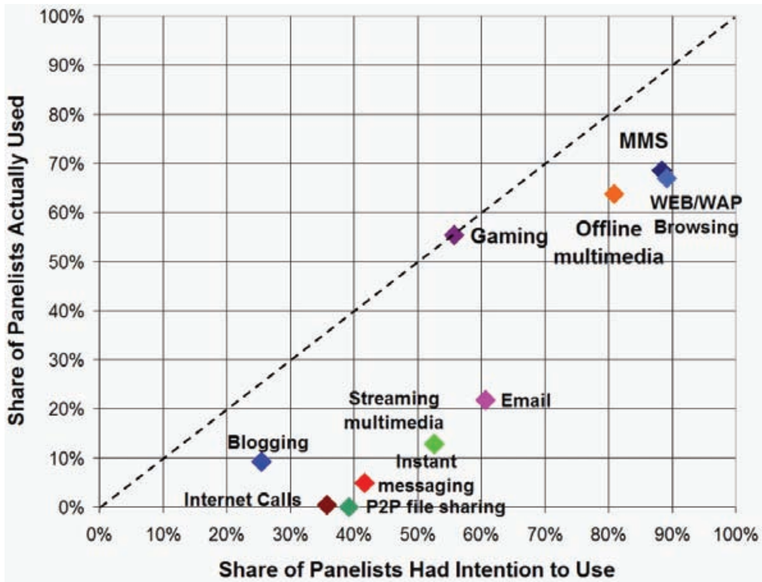


data. Figure 2 projects the identified key smartphone services in two dimensions: the share of panelists who intended to use the service before the panel (*intention index*), and the share of panelists who actually used them (*usage index*). The projected diagonal line depicts the points in which the service's all users who have intentions to use the service actually used the service during the panel. In some cases it, however, might be that some users who did not intend to use the service actually used it anyways. The services that fall off the line have some bottlenecks in the adoption process because not all of the interested panelists used them. Conversely, those services that are above the line have been adopted by more panelists than who actually intended to do so.

WEB/WAP browsing, MMS, offline multimedia and gaming have experienced usage approximately according to intentions. In other words, approximately those who intended to use the services also were able to do so. Therefore these services can be considered as sort of successes in this study. More interestingly, there are many services that lie relatively far away from the diagonal line. These include e.g. instant messaging, mobile email (both embedded email clients and webmail included) and streaming multimedia. People had intentions to use these services, but for a reason or another usage did not realize. Some bottlenecks for adoption remain for some services, whereas for some services bottlenecks are significantly lower. All services that received high usage indices (located in the upper right hand corner) have been available in the market for longer than those that received low usage indices. Clearly service maturity reflects in the results.

Figure 3 demonstrates how intentions correlate with service adoption. Those who stated they have strong interest towards the service had significantly higher adoption rates. Strong correlation between intentions and the consequent adoption process exists. As was found, WEB/WAP browsing, MMS, gaming and offline multimedia playback catch a wide domain of those users who originally had intentions towards the particular service. On the other hand, only 60% of those who had a very strong interest to use mobile email and 33% of those who had a very strong interest to use various streaming multimedia services actually did use these services. There are many arguments to explain why certain services experience good adoption whereas some others do not. Some of

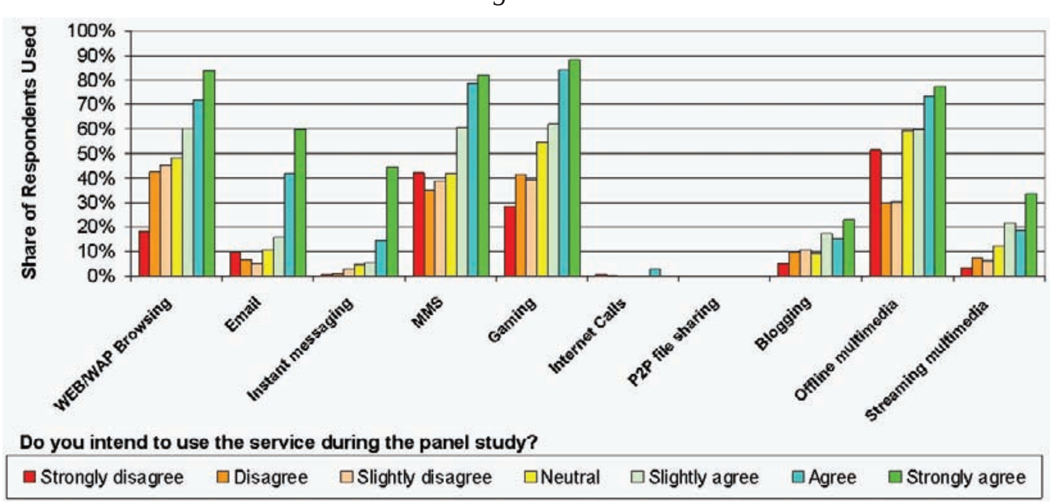
Figure 2. User intentions and actual use of services



these reasons include the technical difficulty to configure/use the service, immaturity of the service, unavailability of necessarily technical help, problems in the marketing / communication functions and suboptimal pricing.

The figures demonstrate that there exists a gap between user intention and actual usage. This is called here as the "adoption gap", as it reflects the problems in moving from intentions to actual usage. This chapter develops a service *adoption index* measuring the extent of this gap. The index can be calculated by measuring the share of panelists who actually used the service out of those who intended to use it. For services that experience little bottlenecks in adoption, this should be close to 100%, whereas for services

Figure 3. User intentions and actual service usage



with significant bottlenecks this should be closer to 0%. Graphically all services located close to the diagonal line have high service adoption indices. The advantage of this index is that it can be used when comparing services with different customer awareness (e.g. different diffusion levels of original intention to use), as effectively the measure is normalized based to the number of people who intended to use the service (instead of all panelists).

In addition, there exists a gap between intended usage (reflecting current short term interest to explore the service) and expected future service diffusion (reflecting future expectations for the service, i.e. is the service going to succeed in the mobile market by extending the "wired" version of the service). A background question was asked with regards to users' expectations of different mobile services succeeding in the mobile market by replacing their use of the service in other platforms (e.g. desktop computers or MP3 players). This is called as an attitude towards the service variable in this research as it reflects positive expectations for the mobile version of the service. Rather than reflecting general attitude it reflects positive long-term perception of the service eventually being used by the panelist herself. *Attitude index* measures the share of panelists having a positive attitude towards the service. Further, *timing index* measures the share of panelists having positive intentions to use the service during the panel (in the short term) out of those panelists having a positive (long-term) attitude towards the service. This variable is very useful, as it normalizes the number of people intending to use the service by the number of people having long-term interest (i.e. attitude) towards the service. Therefore the timing index can be used in evaluating whether the market is ready to adopt a particular service right now, given that positive attitudes and therefore eventual latent demand for the service exist. The timing index is high if a relatively high share of interested panelists are likely to adopt a given service in the short-term, and low if service adoption is for a reason or another (e.g. unavailability of the service) not acute right now.

Table 1. Measured service indices

	Attitude Index	Intention Index	Timing Index	Usage Index	Adoption Index	
WEB/WAP Browsing	73%	88%	92%	69%	72%	
Email	87%	61%	65%	22%	31%	
Instant messaging	79%	42%	42%	5%	10%	
MMS	94%	89%	90%	67%	70%	
Gaming	68%	56%	58%	55%	71%	
Internet Calls	83%	36%	34%	0%	0%	
P2P file sharing	63%	39%	41%	0%	0%	
Blogging	76%	25%	26%	9%	13%	
Offline multimedia	95%	81%	82%	64%	70%	
Streaming multimedia	85%	53%	52%	13%	19%	

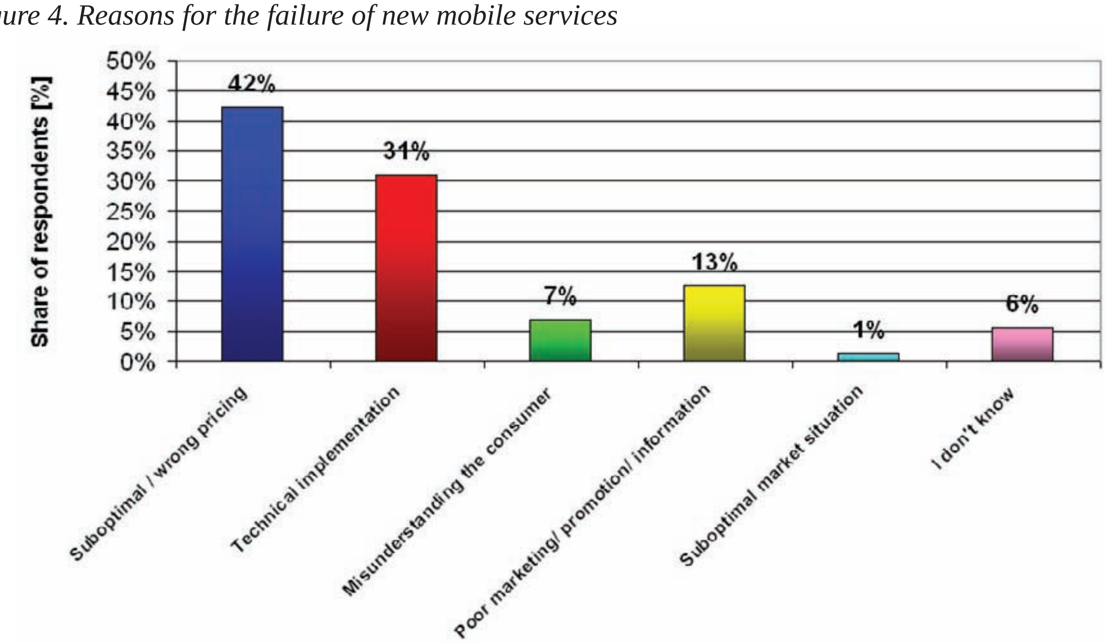
Attitude index	Share of panelist thinking that the mobile service has replaced / will replace corresponding existing use of the service (i.e. computers)
Intention index	Share of panelists having an intention to use the service in the near future
Timing index	Share of panelists intending to use the service in the near future out of those having a (long-term) positive attitude
Usage index	Share of panelists having used the service during the panel study
Adoption index	Share of panelists having actually used the service out of those who intended to do so

The Table 1 summarizes the calculated indices for the identified services. By looking at the attitude index, it can be said that people have positive attitudes (future expectations for mobile services replacing their use of corresponding services with other devices) towards most services, the most so with regards to offline multimedia playback, MMS (replacing corresponding multimedia messaging clients in desktop computers, such as some IM clients) and mobile email. The second column (intention index), however, tells that most people have short-term interest to use only browsing, MMS, games and offline multimedia services. Indeed, the timing index is higher than 50% for these services. They are likely to generate usage right now. Some other services (having high attitude indices) might be successes further in the future according to the panelists, an example being e.g. mobile blogging which has a rather positive attitude index but low intention and consequently timing index. Also mobile VoIP (Internet calls) have promising prospects, as many people have positive long-term attitudes towards the service though no short-term interest (intention) to use the service exists. The timing index therefore reflects current propensity of taking service into use, given that there is a positive attitude towards the service. The fourth column projects the actual share of panelists having used the service during the panel, and finally adoption index in the fifth column states that only offline multimedia, browsing, MMS and gaming have been adopted without significant bottlenecks. All of these five indices can be measured over time. Longitudinal comparisons would provide visualizations on the development of services over their individual adoption curves.

The panelists themselves have opinions with regards to the key bottlenecks for new mobile services not to actually succeed in the market. All together 71 panelists were asked randomly (utilizing the pop-up questionnaire functionality in the research platform) why they think new mobile services do not survive in the market. Each respondent could pick only one answer, and the results are projected in the Figure 4.

According to the panelists it is still pricing that has the most influence on actual usage. In other words, even though people have interests, too high a price level or suboptimal pricing structure might prevent them from using a particular service. The second most important thing the

Figure 4. Reasons for the failure of new mobile services



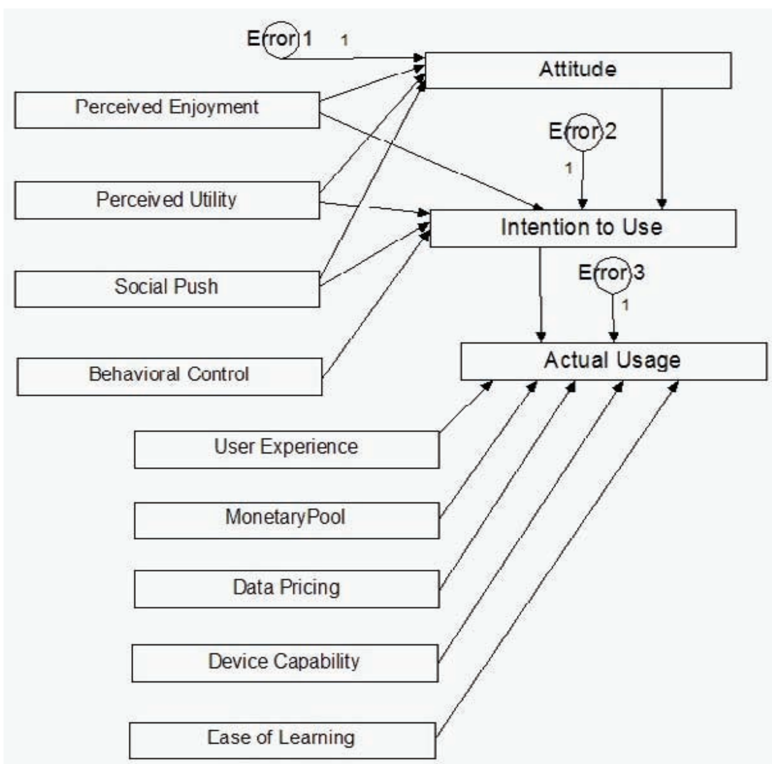
respondents identified as a bottleneck is technical implementation. Many services (according to the panelists particularly handset-embedded mobile email) are simply too difficult to use or configure, and this serves as a bottleneck. The research now turns to studying the adoption process in a more detailed manner by utilizing path modeling and comprehensive background data.

3.3. Path Analysis Model

Path analysis is used as a tool in verifying the theoretical hypotheses of this chapter. Path analysis is an extension of multiple regression. Wright (1921; 1934; 1960) was first to utilize path analysis in empirically studying direct and indirect effects of theoretical models. In essence, path analysis extends multiple regression by including a number of equations instead of only one (Schumacker & Lomax 2004).

Pedersen and Thorbjørnsen (2003) is used as a basis for the theoretical model developed here. Therefore factors perceived enjoyment, perceived usefulness, social push (called in other contexts as subjective norm) and behavioral control (i.e. perceived ease of use) are included to have either direct or indirect (mediated by attitude) effect on intention to use. The perceived expressiveness variable is not included in this model to keep it simpler. After all, it has not been included in all of the earlier TAM models. For the attitude variable a question was asked with regards to a particular service having already replaced or replacing potentially in the future the existing use of the service with e.g. computers or MP3 players. If people have a positive attitude towards the service, they answer positively to this question. Each factor is measured by a single question, listed in Appendix A. Due to limited data (and consequent unobserved uncertainty) the

Figure 5. Developed theoretical path model of mobile service adoption



developed path model does not work that well by itself. However, the path models estimated for each of the services can still be compared against each other in analyzing the role of each factor in the context of a cross-service study. The original TAM model assumes that demographics (such as gender and age) and many other background variables are totally mediated out by the included independent factors in the extended TAM model. This simplifies the analysis.

Because of the variety of data available not only related to background variables but also on actual usage, the theoretical model is extended to include a factor reflecting actual usage (usage frequency of the service). Actual use is depicted to depend on intention to use, technical capability of the phone (the handset features serve either as a bottleneck or enabler), data pricing (lower marginal data prices should push services that generate data charging records), monetary pool available for smartphone usage (more monetary resources should have a positive impact on service usage that generate charging records) and user experience (more experienced users should be more likely to be capable of using the service). The technical capability of the phone is a boolean variable taking value 1 if the handset is based on Nokia S60 3rd edition platform (all new Nokia smartphones) and 0 otherwise (older Nokia S60 handset). Data pricing takes values 0 (no data plan, expensive usage-based charging), 1 (block-priced data plan, relatively cheaper to use data services) and 2 (flat-rate, unlimited usage with zero marginal cost). Monetary pool is proxied through ABPU (average billing per user) rates of subscribers, hypothesizing that high average bills of the past reflect the user's mobile service budget. User experience is here communicated through estimated average minutes of smartphone usage per day (that has strong correlation to user's experience according to a smaller sample of data).

One more variable is included to reflect a positive perceived easiness of learning and ease of getting help (i.e. facilitating conditions in actually using the service) for the service. This is projected to have a direct impact on actual use. The behavioral control variable introduced earlier better reflects the expected difficulty of use thus having implications in the initial intention to use, whereas the ease of learning component is hypothesized to be relevant if the initial interest towards the service exists and the problem is how to actually use the service given positive intentions.

The models are estimated with AMOS (add-on package for SEM in SPSS), using the maximum likelihood approach. Because there are missing data points (not all people answered the questionnaires), only panelists with full data are included. All in all, 426 panelists are included in the path estimation.

3.4. Results of Path Analysis

The estimated path analysis diagrams for the chosen services are depicted in Appendix B. The figures include path coefficients (above the paths) and for dependent variables the share of variance explained by the model (above the box representing the variable). Some model fit indices are reported, too. First, a fit index CMIN/DF is included - the relative chi-square that represents how much the fit of data has been reduced because one or more paths have been dropped from the full model. Generally if this index is more than 3, too many paths have been dropped. The second included fit index is GFI (goodness of fit index). This tells how much of the variance in the sample variance-covariance matrix is accounted for by the model. Generally this should be above 0.9. Finally RMSEA (root mean square error of approximation) estimates lack of fit if compared to the full model. 0.08 or below is considered generally as adequate fit.

By looking at the results in Appendix B, the first observation is that the fit indices are quite poor. In other words, the developed model is perhaps not the best one to explain the involved complex relationships. GFI indices range from 0.82 (IM and email) to 0.86 (browsing). The second observation is that the model quite poorly explains the dependent variables under interest. The variance explained for the intention to use variable range from 0.05 (blogging) to 0.24 (gaming and MMS). The variance explained for the actual usage variable range from 0.01 (blogging) to 0.27 (gaming). Lots of uncertainty (e.g. non-internalized variables) exists that would explain intentions and actual usage, and the model therefore cannot depict all of the variance in the dependent variables. Almost all of the variables included in the framework model the demand-side, and therefore the missing (explanatory) variables are probably related to supply-side factors (e.g. service push, pricing and availability).

Without going into the details of each service (the estimated path models can be found from Appendix B), Table 2 summarizes the statistically significant path coefficients for each service.

It is surprising that perceived enjoyment has very strong impact on attitude for all services. People who think the enjoyment value is high (for any service) have high expectations that they are going to replace existing use of the service with the mobile version of the service (this is here called considered positive attitude). Perceived enjoyment positively (and statistically significantly) drives intentions directly, too. It can be generalized that the advanced mobile service market in Finland is building on expected hedonic benefits even in services which have been generally considered business-oriented (email).

Further, utilitarian benefits are important, too. The impact towards attitude is positive for all services expect for streaming multimedia, though the relationships are not as strong as with hedonic benefits. Interestingly the positive forward-looking attitude towards mobile instant messaging is strongly pushed by perceived utilitarian benefits. This confirms that instant messaging is a viable communication channel rather than a service meant for entertainment purposes solely. The direct impact of perceived utility on intentions is not that strong as the impact of perceived enjoyment (hedonic benefits), but the model nevertheless confirms that the intentions to use browsing, MMS, email (strong effect), blogging and IM are at least partly driven by utility-related expectations. Expectedly perceived utility does not drive offline multimedia playback, gaming or streaming multimedia usage intentions, because they are quite entertainment-oriented by nature.

Social push is all about the impact of supportive/driving social context with regards to usage attitude and/or intentions. Social push has positive impacts particularly on email and instant messaging. This can be explained by the fact that these services are currently available, and the marginal impact of somebody in the close social neighborhood helping or recommending the service might be significant to drive attitude towards possible future use. The direct effect of the social push on intention to use is not very strong for any of the other services, communicating the fact that these early-adopter users are quite independent.

Behavioral control measures the user perception of herself being capable of using the service and to overcome technical challenges. Particularly for browsing and gaming this factor is a significant driver of intentions, whereas for most other services it has a positive though less significant effect. The statistical significance for the path originating from behavior control is generally not that strong. This might be due to the fact that Hejden et al. (2005) hypothesized that for hedonic services behavioral control does not have that strong an impact. For MMS there is no relation at all, as MMS is already well integrated into smartphones and thus there is little variance in end-user behavioral control.

The TAM model suggests that attitude towards the service should drive intentions. The attitude variable included in this research is geared towards current or future role of the mobile service possibly replacing existing use of the service with other devices (e.g. laptops or MP3 players). Therefore

Table 2. Statistical significance of estimated paths

Path	Browsing	MMS	Email	IM	Offline multimedia	Streaming multimedia	Gaming	Blogging
Perceived Enjoyment -> Attitude	***	***	***	***	***	***	***	***
Perceived Utility -> Attitude	**	**	**	***	**		***	**
Social Push -> Attitude	*		**	**	*			
Behavioral Control -> Intention to Use	**		*	*	*	*	***	*
Perceived Enjoyment -> Intention to Use	***	***	***	***	***	***	***	
Perceived Utility -> Intention to Use	*	*	***	*				***
Social Push -> Intention to Use				*	*			
Attitude -> Intention to Use			*		**			
User Experience -> Usage	***	**			***		***	
Monetary Pool -> Usage		**			*			
Data Pricing -> Usage	***		**					
Device Capability -> Usage			***		***	**		
Ease of Learning -> Usage								
Intention to Use -> Usage	***	***	***	***	***	**	***	
	***	p≤0.001		(positive relation)				
	**	p≤0.01		(positive relation)				
	*	p≤0.05		(positive relation)				
	*	p≤0.05		(negative relation)				

this variable is not reflecting general attitude in the purest sense. There is a statistically significant relation between attitude and intention only for mobile email. Positive mobile email attitude levels correlate strongly with positive intentions. People who think mobile email is or will be an important part of their smartphone usage in the future also have strong short-term intentions of using the service. For many other services the effect of attitude on intention is statistically insignificant. This can be explained by the fact that people generally perceive that these new advanced services are unavailable / technically poor, and therefore they lack short-term intentions of use though there is a positive general attitude. Another argument is relatively stronger effects of direct effects of background variables. In other words, rather than for many background factors having an indirect impact via positive attitude that the service would replace something (having a flavor that something else is sacrificed), for example perceived enjoyment or utility directly drive intentions and these relations are very strong.

When moving to actual use of services, it becomes clear that the variables hypothesized to have direct effect on realized usage are important. User experience (proxy through average usage minutes per day) has statistically significant relationships with actual usage of browsing, MMS, offline multimedia playback and gaming. These are services users have probably used already earlier with their smartphones, and therefore experience variable strongly drives usage. For some very new services such as email, IM or streaming the experience variable does not have a statistically significant effect. This might be because earlier experience relates to services that have been in the market already for some time (on which experience can accumulate), whereas experience does not help with regards to services that are just about to hit the market or are available later in the future.

Monetary pool has a strong effect on MMS usage, which has strong correlation with other mature service (such as voice and SMS) usage. In other words, higher consumption levels on mobile telephony do not drive any of the truly new service categories. First, some services are simply free and they do not depend on consumption levels or available monetary pool but on e.g. perceived enjoyment instead. Second, still most of the ABPU (average billing per user) rate consists of voice and SMS service charging, and therefore it does not reflect data service prices/consumption that well. All of the services that require network connectivity in this research are based on IP connectivity and are therefore independent of circuit-switched service pricing.

Data pricing (the lower marginal cost of using data services) has the most significant relationship expectedly with browsing usage. Browsing is currently the most important data (Internet) service. There is also weaker (but still statistically significant) effect on email usage. However, for other data services data pricing does not have an effect. This is partly due to little actual usage realized for other than browsing data service.

Interestingly device capability has the biggest influence on email and multimedia functionalities. The impact on multimedia is not a surprise, as if nothing else, new handsets include at least better multimedia functionalities (camera chips, memory capacity, MP3 player functionalities, better displays for movie/video playback) than older ones. The marginal email usage is coming from webmail oriented email usage. This is partly driven by better browsers included in newer handsets together with new high-resolution displays. This explains the path from device capability to email usage.

Ease of learning variable does not have an impact on actual usage at all. These early-adopter users, after all, all know that there is help available for example in the Internet. On the other hand, they also tend to play around with the handset by themselves rather than think that they need somebody else for help, not to talk about manuals that are shipped with new handsets.

Finally, the most important variable in the model (intention to use) has statistically significant effects on all service usage expect for blogging. As concluded already earlier with descriptive statistics, people who have strong intentions to use the service tend to adopt the service (use it in practice) with a higher probability than those who do not have intentions. Blogging has not been used by that many, and therefore intentions cannot correlate with usage. Most blogging usage comes actually from Nokia's preinstalled Lifeblog application. Because of this it can be that those who explored blogging merely randomly launched Lifeblog, even though they did not have clear intentions. All in all, usage intentions are still the most important driver of advanced mobile service usage, therefore suggesting that the needs of end-users are the key issue in bringing new mobile services to the market instead of push-centric service introduction strategies.

A couple of reasons exist for the general poor fit of the estimated models. First, these are advanced services, and there are probably many random factors that lead into a usage event. The service must be in the market available for panelists and the technical performance/accessibility has to be in order. These are some of the variables the model cannot internalize, and this explains the poor share of variance explained for usage-level variables. Similarly intentions originate from many other factors (in advanced services) than those included in the model. For example, promotional events and availability of the service certainly do affect the short-term intention to use a service. These variables are not included in the model. In many cases these reasons explain the poor fit of the model. Indicative results suggest that the depicted theoretical model is not the optimal one. For example mobile blogging is a rather new service on which customers have very little or limited knowledge on. For this service most paths are insignificant and model fit is very poor. In addition, it is difficult to explain the variance of dependent variables if these variables have no or very little variance. Gaming is a well-known service and the market-level factors (e.g. operator competition, network performance, pricing, promotional events) have less impact. Therefore the theoretical adoption model receives much better results for gaming than for blogging, for example.

4 CONCLUSION

The chapter utilized a newly developed handset-based mobile service research platform in deploying questionnaires and measuring actual usage of mobile services. This study differs from other handset-based research in its specific focus on the adoption process and associated measurements, utilizing also questionnaires. Descriptive statistics reveal that for most services not all mobile service demand (i.e. usage intentions) is fulfilled, i.e. not all panelists interested in services are actually using them. However, those strongly intending to use the service, ceteris paribus, expectedly have higher probabilities of adopting the service in practice.

The chapter developed five indices that can be measured over time and therefore utilized in longitudinal analysis. These indices communicate the general attitude towards mobile services (attitude index), short-term intentions of usage (intention index), the current propensity of taking services into use (timing index), the extent of service usage (usage index) and the probability of using the service given intentions exist (adoption index). These indices communicate that panelists on average have positive attitudes towards mobile versions of the services included in the study, but only matured mobile services (introduction more than a year ago) experience positive short-term intentions of usage and consequent adoption. Different services are clearly experiencing different phases of their life-cycle at this point of time. The developed indices can be utilized in projecting the phase of diffusion for each service emerging in the market.

The estimated path models reveal that perceived hedonic (enjoyment) benefit is the strongest driver of both attitude and intention towards the service. This is in contrast to Pedersen and Thorbjørnsen (2003) who found that utility-centric benefits drive intention to use mobile services. In the estimated models of this chapter perceived utility drives intention to use only in the case of mobile email. In general, many of the newly developed mobile services should be considered as generating hedonic rather than business/utilitarian value to end-users. The social setting around panelists or expected technical difficulties do not explain intentions to use. This is most likely due to panelists in the dataset being early-adopters (quite independent and technically advanced users). User experience with smartphones explains actual use of services that have already been in the market for some time, but experience of the user does not explain the adoption of the most recent services. Pricing of data traffic has a strong effect only on browsing and email as they are the most visible mobile Internet services currently. Multimedia services, on the other hand, benefit from higher capability of handsets. The most important thing driving actual use of services, however, is the intention of end-users to use the service, and therefore user needs should be looked upon in the future instead of technology push strategies.

The estimated path models do not have very good fit. Instead of using standardized sets of questions and structural equation modeling in verifying the theoretical model, a more simplified path modeling exercise was chosen due to limitations in the panel study implementation. Improvement of the theoretical model remains as a future work task in addition to the acquisition of more comprehensive dataset supporting not only path but also measurement part of the SEM analysis. One of the key conclusions of this chapter is, however, that the internalized variables of the estimated adoption model poorly explain the variance observed in intention and usage-level variables. Therefore most advanced mobile services have special characteristics and external sources of variance that were not observed in this analysis. These likely originate from the supply side of the market. Nevertheless, the demonstrated fitting of empirical data with path analysis was found to be an efficient method to compare the role of different variables in explaining the adoption of emerging mobile services.

Some mobile services are truly successful in the market, whereas some others are hyped quite a lot but few people actually adopt them. A need exists to acquire data on user opinions and service usage in order to understand the actual dynamics of mobile service adoption. This understanding is valuable in e.g. better commercializing mobile services and in improving mobile customer relationship management processes (for mobile CRM see Liljander et al. 2007). This chapter showed that new handset-based data can help in better understanding the mobile service adoption process. The research differed from other handset-based research in focusing on the modeling of the adoption process, and combining efficiently questionnaire data with usage data. The chapter introduced many new indices that reflect the different sides of the adoption process, and additionally demonstrated that different services can be compared against each other with the handset-based research approach. Future handset-based research should deploy wider questionnaires to make the statistical procedure of path modeling more reliable. In particular, the measurement part of the analysis should be deployed before estimating the structural part of the path model. In addition, a wider perspective towards the adoption process should be taken. For example, potential long-term usage should be better modeled against potential short-term usage, and reasons why explorative usage sometimes does not lead to sustainable and repetitive usage should be looked upon.

5 REFERENCES

Aarnio, A., Enkenberg, A., Heikkilä, J., & Hirvola, S. (2002) Adoption and Use of Mobile Serivces - Empirical Evidence from a Finnish Survey. *Proceeding of the 35th Hawaii International Conference on System Sciences 2002* (1454-1463).

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Ed.), *Action Control: From Cognition to Behavior.* New York: Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organization Behavior and Human Decision Processes, 50,* 179-211.

Bakalis, S., Abeln, M., & Mante-Meijer, E. (1998) *Adoption and use of mobile telephony in Europe*. Cost 248 Report. Telia, Farsta, Sweden.

Burton-Jones, A., & Hubona, G. S. (2005). Individual Difference and Usage Behavior: Revisiting a Technology Acceptance Model Assumption. *The DATA BASE for Advances in Information Systems - Spring 2005,36*(2), 75.

Christensen, C. M. (1997). *The Innovator's Dilemma*. Harvard Business School Press.

Chuang, M. C., Chang, C. C., & Hsu, S. H. (2001). Perceptual factors underlying user preferences toward product form of mobile phones. *International Journal of Industrial Ergonomics, 27,* 247-258.

Clark, D. D., Lehr, W., Bauer, S. J., Faratin, P., Sami, R., & Wroclawski, J. (2006). Overlay Networks and Future of the Internet. *Journal of Communications and Strategies, 3*(63), 1-21.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science, 35, 982-1003.*

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*, 319-340.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading MA: Addison-Wesley.

Flanagin, A. J., & Metzger, M. J. (2001). Internet use in the contemporary media environment. *Human Communication Research, 27,* 153-181.

Funk, J. L. (2004). *Mobile Disruption - the technologies and applications driving the mobile internet.* Wiley.

Gao, M., Hyytinen, A., & Toivanen, O. (2006). Demand for and Pricing of Mobile Internet: Evidence from a Real-World Pricing Experiment. Helsinki Center of Economic Research Discussion Paper No. 122.

Green, N., Harper, R. H. R., Murtagh, G., & Cooper, G. (2001). Configuring the mobile user: sociological and industry views. *Personal and Ubiquitous Computing*, *5*, 146-156.

Haddon, L. (2001). *Domestication and mobile telephony.* Presented at the conference "Machines that Become Us", Rutgers University, April 18-19, New Jersey.

Heijden van der, H., Ogertschnig, M., & van der Gaast, L. (2005). *Effects of Context Relevance and Perceived Risk on User Acceptance of Mobile Information Services.* ECIS 2005.

Heinonen, K., & Pura, M. (2006). Classifying Mobile Services. *Presented at Helsinki Mobility Roundtable*, 1-2 June 2006, Helsinki.

Höflich, J. R., & Rössler, P. (2001). Mobile schriftliche Kommunikation oder: E-Mail für das Handy. *Medien & Kommunikationswissenschaft, 49,* 437-461.

Karlson, B., Bri, A., Link, J., Lönnqvist, P., & Norling, C. (2003). *Wireless Foresight: Scenarios of the Mobile World in 2015*. John Wiley and Sons.

Kearney, A. T. (2001). *Satisfying the experienced on-line shopper: Global e-shopping survey.* A.T. Kearney Ltd., Inc. London.

Kim, J. (2001). Analyzing mobile Internet users. *Presented at the CHI 2001 Workshop on mobile communications.*

Leung, L., & Wei, R. (2000). More than just talk on the move: Uses and gratifications of the cellular phone. *J&MC Quarterly , 77,* 308-320.

Liljander, V., Polsa, P., & Forsberg, K. (2007). Do Mobile CRM Services Appeal to Loyalty Program Customers? *International Journal of E-Business Research, 3*(2), 24.

Ling, R., & Yttri, B. (2002). Hyper-coordination via mobile phone in Norway. In J. E. Katz & M. Aakhus (Eds.), *Perpetual contact*. New York: Cambridge University Press.

Ling, R. (2001). *"It is 'in.' It doesn't matter if you need it or not, just that you have it.":* Fashion and the domestication of the mobile telephone among teens in Norway. Oslo: Working Paper, Telenor R&D, Norway.

Mallat, N. (2006a). *Consumer and merchant adoption of mobile payments.* Doctoral dissertation. Helsinki School of Economics 2006.

Mallat, N. (2006b). *Exploring Consumer Adoption of Mobile Payments.* Presented at Helsinki Mobility Roundtable, 1-2 June 2006, Helsinki.

Meister, D. B., & Compeau, D. R. (2002). Infusion of Innovation Adoption: An Individual Perspective. *Proceedings of the ASAC, Winnipeg, Manitoba*.

Minges, M. (2005). Is the Internet Mobile? Measurements from the Asia-Pacific region. *Telecommunications Policy 29 (2005), 113–125.*

Nokia Telecommunications. (1999). *The Demand for Mobile Value Added Services: A Market Study*, available at http://www.nokia.com/press/background/pdf/study-vas.pdf, 1999.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005b). Explaining intention to use mobile chat services: Moderating effects of gender. *Journal of Consumer Marketing, 22*(5), 247-256.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005a). Intentions to Use Mobile Services: Antecedents and Cross-Service Comparisons. J*ournal of the Academy of Marketing Science*, *33*(3), 330-346.

Pedersen, P. E., & Thorbjørnsen, H. (2003). Adoption of Mobile Services. *Model development and Cross-Service study.* Norwegian School of Economics and Business Administration.

Pedersen, P. E. (2001). An adoption framework for mobile commerce. *Proceedings of the 1st. IFIP Conference of E-Commerce*, Zürich, Switzerland, October 3-5, 2001.

Pedersen, P. E. (2005). Adoption of mobile Internet services: An exploratory study of mobile commerce early adopters. *Journal of Organizational Computing and Electronic Commerce*, *15*(3), 203-221

Pedersen, P. E., Methlie, L. B., & Thorbjørnsen, H. (2001). *Understanding mobile commerce end-user adoption: A triangulation perspective and suggestions for an exploratory service evaluation framework* . Presented at HICSS-35, Hawaii, US, Jan 7-10, 2002

Rask, M., & Dholakia, N. (2001). Next to the customer's hart and wallet: Frameworks for exploring the emerging m-commerce arena. *AMA Winter Marketing Educator's Conference*, *12*, 372-378.

Rogers, E. M. (1962). *Diffusion of Innovation*. New York: The Free Press.

Rogers, E. M. (1995). *Diffusion of innovations.* (4th edition). New York: The Free Press.

Saarikoski, V. (2006). *The Odyssey of the Mobile Inter*net. Doctoral Dissertation. University of Oulu, Finland.

Schumacker, R. E., & Lomax R. G. (2004). *A Beginner's Guide to Structural Equation Modeling.* 2nd Edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Sigurdson, J. (2001). WAP OFF - Origin, Failure and Future. *Prepared for the Japanese-European Technology Studies (JETS)*, October 9, 2001, University of Edinburgh.

Skog, B. (2000). The mobile phone as symbolic capital in teenage cultures. (in Norwegian). In *R. Ling and K. Thrane (Eds.), Social consequences of mobile telephony* (in Norwegian). *Proceedings from a seminar on society, children, and mobile telephony*, Telenor FoU - Notat nr. 38/2000.

Skog, B. (2002). Mobiles and the Norwegian teen: identity, gender and class. In J. E. Katz and M. Aakhus (Eds.), *Perpetual contact*. New York: Cambridge University Press.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, *6*, 144-176.

Tseng, F. C., Teng, C. I., & Chiang, D. M. (2007). Delivering Superior Customer Perceived Value in the Context of Network Effects. *International Journal of E-Business Research*, *3*(1), 41.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186-204.

Verkasalo H. (2005). *Handset-Based Monitoring of Mobile Customer Behavior.* Master's Thesis

Series. Networking Laboratory. Department of Electrical and Telecommunications Engineering. Helsinki University of Technology.

Verkasalo, H., & Hämmäinen, H. (2006). Handset-Based Monitoring of Mobile Subscribers. *Presented at Mobility Roundtable*, 1-2 June, 2006, Helsinki, Finland.

Verkasalo, H., & Hämmäinen, H. (2007). A Handset-Based Platform for Measuring Mobile Service Usage. *INFO: The Journal of Policy, Regulation and Strategy, 9*(1), 2007.

Verkasalo, H. (2006a). Mobile Data Service Evolution - Empirical Implications from Europe and the USA. In *3rd International CICT Conference*, November 30 - December 1, 2006, Denmark, 2006.

Verkasalo, H. (2006b). Empirical Observations on the Emergence of Mobile Multimedia Services in the U.S. and Europe. In *The 5th International Conference on Mobile and Ubiquitous Multimedia*. December 4-6, 2006, Stanford University, California, 2006.

Verkasalo, H. (2007a). *A Cross-Country Comparison of Mobile Service and Handset Usage*. Licentiate's thesis, Helsinki University of Technology, Networking Laboratory, Finland.

Verkasalo, H. (2007b). Empirical Insights on the Evolution of the Finnish Mobile Market. Presented at Conference on Telecommunication Techno-Economics (CTTE) 2007, 14-15 June 2007, Helsinki, Finland.

Verkasalo, H. (2007c). Empirical Findings on the Mobile Internet and E-Commerce. *Presented at 20th Bled e Conference: eMergence: Merging and Emerging Technologies, Processes, and Institutions.* Bled, Slovenia, June 4 - 6, 2007.

Verkasalo, H. (2007d). Contextual Usage-Level Analysis of Mobile Services. *Accepted for publication at The 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS 2007).* August 6-10, 2007 - Philadelphia, PA, USA.

Vesa, J. (2005). *Mobile Services in the Networked Economy.* Hershey, PA, USA: IRM Press

Wong, C. C., & Hiew, P. L. (2005). Correlations between factors affecting the diffusion of mobile entertainment in Malaysia. *ICEC 2005*, 615-621.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557-585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161-215.

Wright, S. (1960). Path coefficients and path regression: Alternative or complementary concepts? *Biometrics, 16*, 189-202.

KEY TERMS

Mobile Internet: Internet for new packet-switched mobile data services. The first mobile data services were hyped quite a lot in the public in the late 90s, but in practice the mobile Internet has not kicked off yet.

Technology Acceptance Model (TAM): TAM models how users come to accept and use a technology. It distinguishes two concepts. First, the perceived usefulness reflects the expected benefits from using a certain technology. Second, the perceived ease of use reflects pretty much the same thing as the perceived behavioral control in the theory of planned behavior.

ENDNOTE

* Reprint of International Journal of E-Business Research (IJEBR) , Volume 4, Issue 3, July-September 2008, pp. 40-63.

APPENDIX A: QUESTIONS ASKED IN THE BACKGROUND QUESTIONNAIRE

Background factors are recoded into the following scale:

ABPU... 1 (small) - 5 (big) (asked before the panel)
Work... 0 (not) - 1 (yes) (asked before the panel)
Age... 0 (young) - 1 (old) (asked before the panel)
Gender... 0 (women) - 1 (men) (asked before the panel)
Device_Capability... 0 (old device = lower capability) - 1 (new device = higher capability) (asked before the panel)
Data_Pricing... 0 (usage-based), 1 (block-priced), 2(flat) (asked before the panel)
Experience (of smartphone usage in the user's own opinion)... 1 = Beginner, 2 = Normal, 3 = Experienced, 4 = Very experienced (asked before the panel)
Panel_Days... Number of days observed active in the panel (derived from usage data)
Handset_Usage_Activity... Number of application activations during an average day (derived from usage data)
Smartphone_Usage_Day... Minutes of smartphone usage on an average day (derived from usage data)

Questions are asked in relation to the following mobile services:

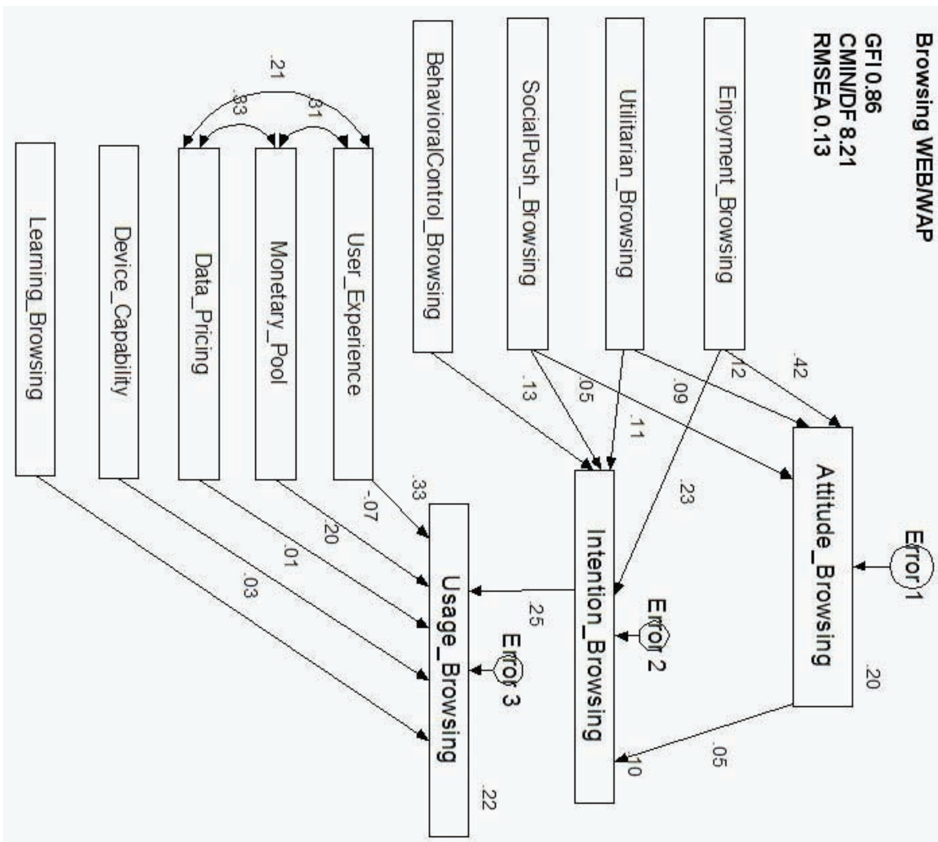
WEB/WAP services, Internet browsing
Email
Instant messaging / chat
MMS messaging
Gaming
Internet telephony
P2P file sharing
User-created content creation, blogging
Streaming/Internet multimedia
Offline multimedia

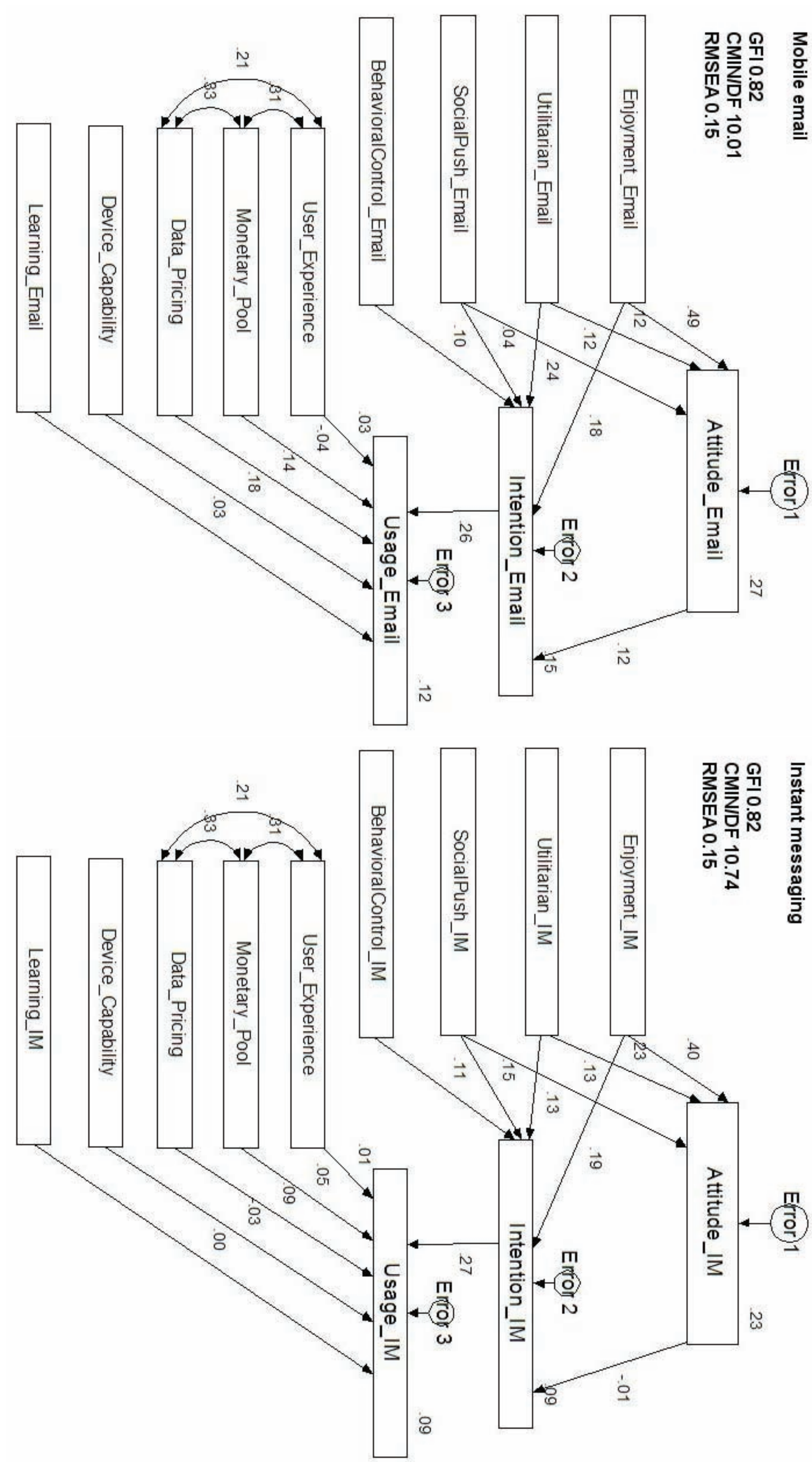
1. I intend to use the following services during the next couple of months… (intention to use) (asked before the panel)
2. Mobile versions of the following services are going to replace / have already replaced the use of those services with other devices (such as PCs, MP3 players, Radio)... (attitude) (asked after the panel)
3. The use of the following services generates enjoyment, pleasure and entertainment to me... (enjoyment value) (asked after the panel)
4. The use of the following services increases my work or study related productivity and performance... (utilitarian value) (asked after the panel)
5. I do not have / would not have significant technical difficulties in using the following services... (behavioral control) (asked after the panel)

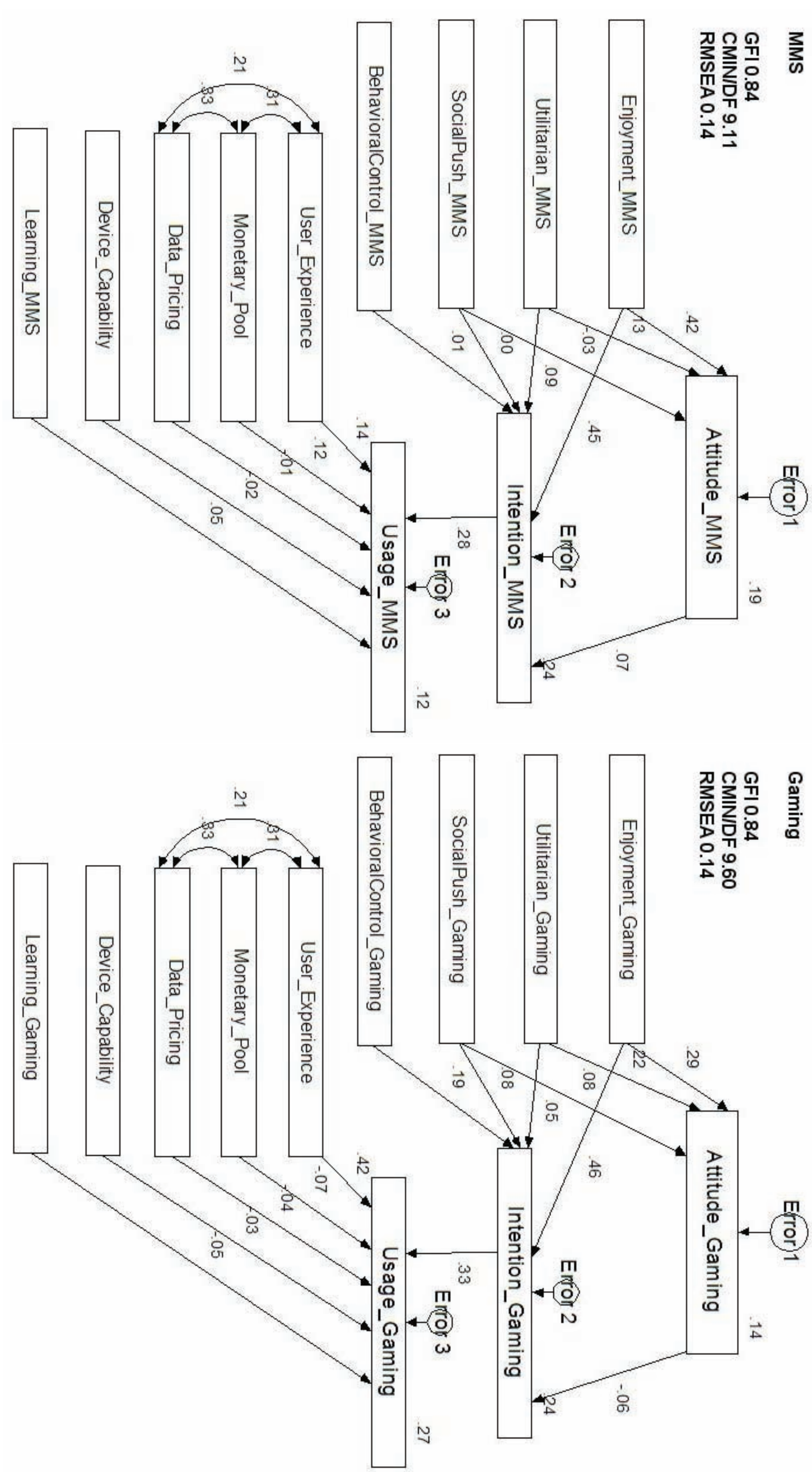
6. It is easy for me to learn and develop my skills in using the following services... (ease of learning) (asked after the panel)
7. People around me (e.g. friends or family) have recommended the use or have helped me in using the following services... (social push) (asked after the panel)

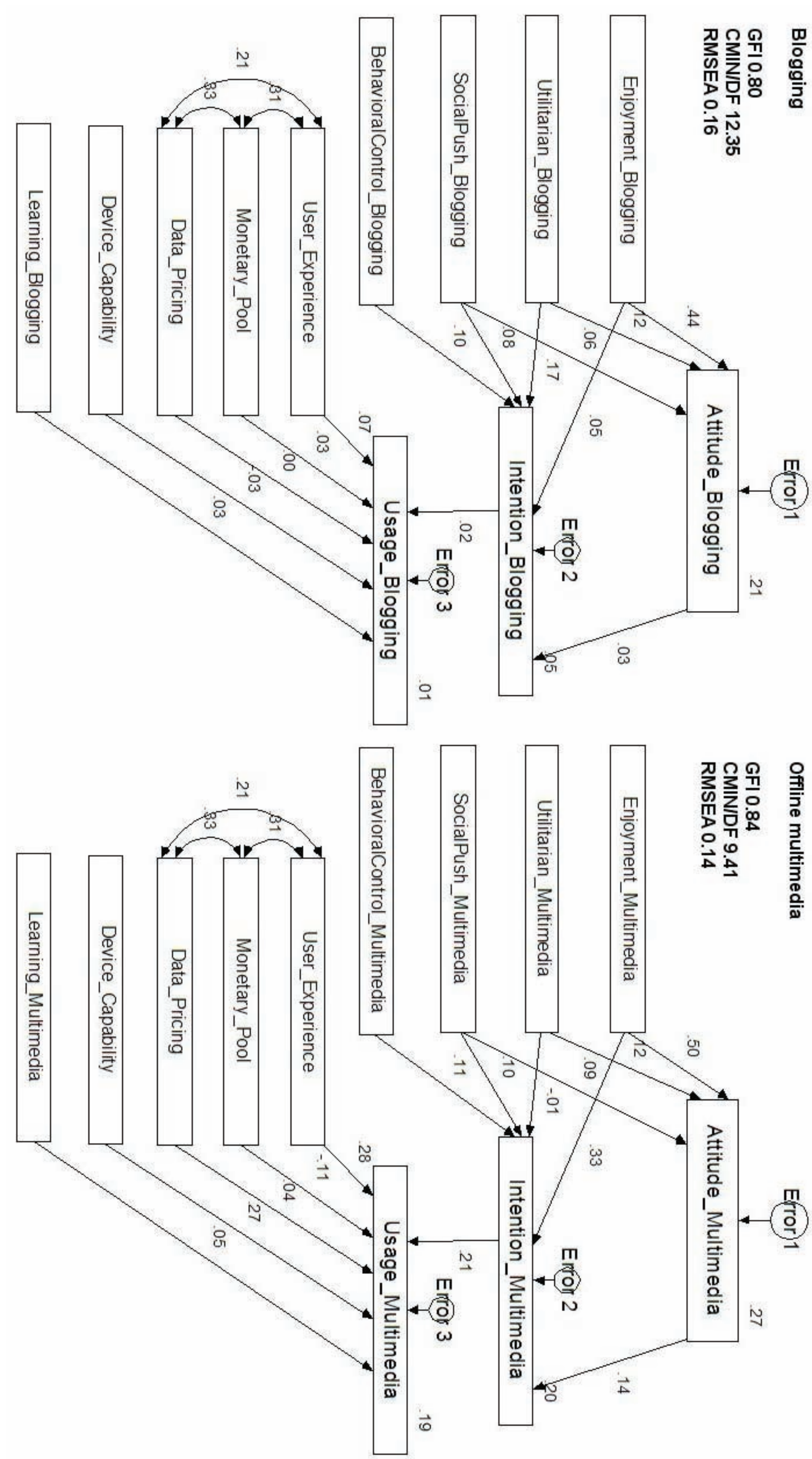
Likert scale answers:
1 = Strongly disagree, 2 = Disagree, 3 = Slightly disagree, 4 = Neutral, 5 = Slightly agree, 6 = Agree, 7 = Strongly agree

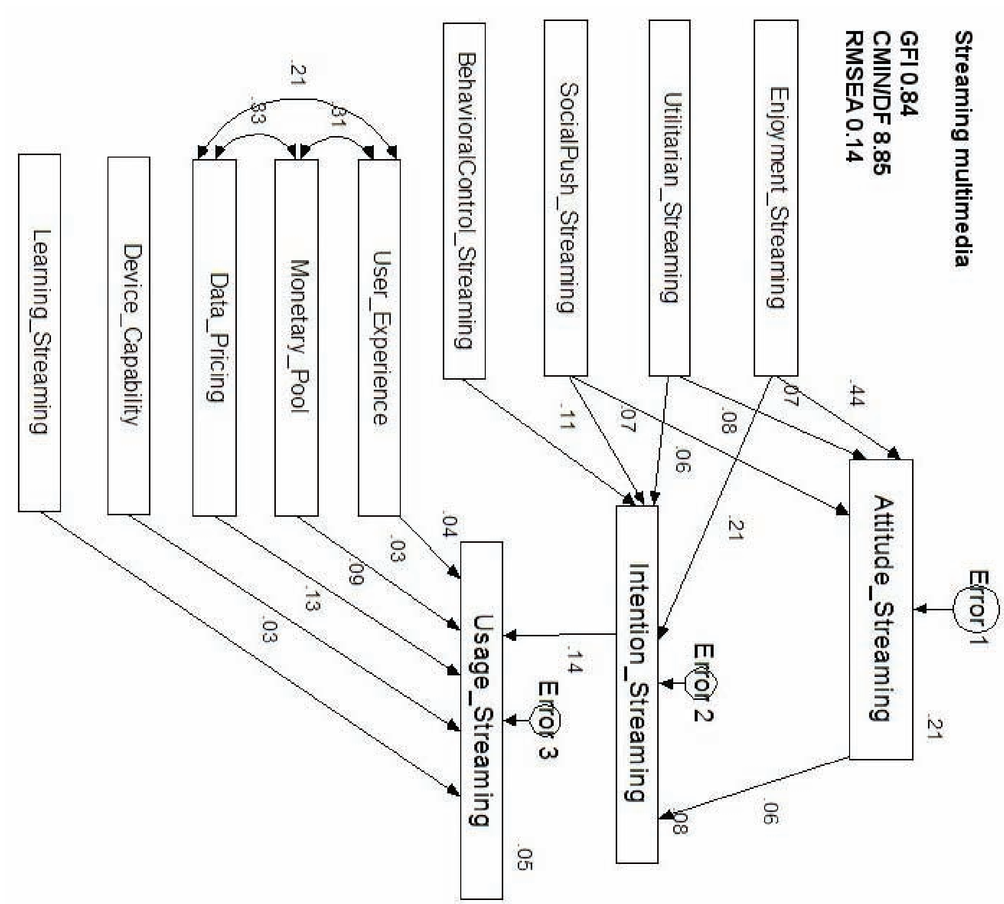
APPENDIX B: RESULTS OF THE PATH ANALYSIS




Mobile email
GFI 0.82
CMIN/DF 10.01
RMSEA 0.15
Enjoyment_Email
Utilitarian_Email
SocialPush_Email
BehavioralControl_Email
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_Email
Attitude_Email
Intention_Email
Usage_Email
Error 1
Error 2
Error 3
Instant messaging
GFI 0.82
CMIN/DF 10.74
RMSEA 0.15
Enjoyment_IM
Utilitarian_IM
SocialPush_IM
BehavioralControl_IM
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_IM
Attitude_IM
Intention_IM
Usage_IM
Error 1
Error 2
Error 3


MMS
GFI 0.84
CMIN/DF 9.11
RMSEA 0.14
Enjoyment_MMS
Utilitarian_MMS
SocialPush_MMS
BehavioralControl_MMS
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_MMS
Attitude_MMS
Intention_MMS
Usage_MMS
Error 1
Error 2
Error 3
Gaming
GFI 0.84
CMIN/DF 9.60
RMSEA 0.14
Enjoyment_Gaming
Utilitarian_Gaming
SocialPush_Gaming
BehavioralControl_Gaming
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_Gaming
Attitude_Gaming
Intention_Gaming
Usage_Gaming
Error 1
Error 2
Error 3


Blogging
GFI 0.80
CMIN/DF 12.35
RMSEA 0.16
Enjoyment_Blogging
Utilitarian_Blogging
SocialPush_Blogging
BehavioralControl_Blogging
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_Blogging
Attitude_Blogging
Intention_Blogging
Usage_Blogging
Error1
Error 2
Error 3
Offline multimedia
GFI 0.84
CMIN/DF 9.41
RMSEA 0.14
Enjoyment_Multimedia
Utilitarian_Multimedia
SocialPush_Multimedia
BehavioralControl_Multimedia
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_Multimedia
Attitude_Multimedia
Intention_Multimedia
Usage_Multimedia
Error1
Error 2
Error 3


Streaming multimedia
GFI 0.84
CMIN/DF 8.85
RMSEA 0.14
Enjoyment_Streaming
Utilitarian_Streaming
SocialPush_Streaming
BehavioralControl_Streaming
User_Experience
Monetary_Pool
Data_Pricing
Device_Capability
Learning_Streaming
Attitude_Streaming
Intention_Streaming
Usage_Streaming
Error1
Error 2
Error 3

Chapter XXII
When Customer Satisfaction Isn't Good Enough:
The Role of Switching Incentives and Barriers Affecting Customer Behavior in Korean Mobile Communications Services

Moon-Koo Kim
Electronics and Telecommunications Research Institute, Korea

Myeong-Cheol Park
Information and Communications University, Korea

Jong-Hyun Park
Electronics and Telecommunications Research Institute, Korea

ABSTRACT

In communications services, the continued competitiveness and growth of a company depends vitally on customer value. In Korea's maturing mobile market currently also going through a period of transition, an intense competition is under way among carriers. As the market nears the point of saturation, carriers are focusing on winning over competitors' subscribers, at the same time on retaining their existing customers. Understanding the factors that influence customers' switching behavior, therefore, is crucial for Korean mobile carriers' quest for successful customer strategies. Customer satisfaction is the widely acknowledged primary determinant of customer behavior. Even so, there are cases in which satisfied customers do not behave according to expectations. Some satisfied customers switch their suppliers, while others stay with their existing suppliers, even if they are dissatisfied with them. In sum, customer satisfaction does not appear to be the sole and the only determinant of customer behavior. This study is an attempt to explain the relationship between customer satisfaction and customer behavior using switching incentives and switching barriers. The role of switching incentives (subsidies toward handset replacement, attractiveness of alternative carriers, etc.) and switching barriers (burden of having to change numbers, burden of losing benefits provided by the current carrier) is investigated empirically using the results of a consumer survey.

INTRODUCTION

When a market matures, the supply of new customers gradually dwindles, and companies are forced to fight harder to retain existing customers and win over the customers of competitors. In a market nearing the point of saturation, retaining existing customers or winning over competitors' customers is a more efficient approach, both in terms of resources and cost, than developing entirely new customers. Suppliers also compete for shares of existing customers in situations where a new product or service replaces existing products or services, rather than creating its own market. Hence, knowledge of customer loyalty and switching behavior is key, both for markets in the maturity phase and those undergoing a transitional phase. Knowing which factors influence customer behavior in a transitional or mature market helps expand our theoretical understanding of consumer behavior. This knowledge can also practically assist businesses with navigating through these market phases.

The Korean mobile communications service industry is exhibiting classical signs of a maturing market, with the number of mobile subscribers remaining stagnant ever since the mid-2000s. The industry is nevertheless in a transitional phase, as 3G HSDPA services are progressively replacing 2G services that have so far dominated the market. The competition for shares of subscribers has become particularly fierce among Korean mobile operators, following the introduction of number portability and unified mobile prefix in 2004. Its three suppliers are faced with the need to both retain their shares of the Korean mobile communications service market and seize the lead in new service areas. They are currently concentrating their organizational resources in efforts to retain customers and win over competitors' customers. As of the end of 2007, over 40% of Korean mobile subscribers have switched their carriers at least once (information published at the website of the Korean Ministry of Information and Communication).

It is the consensus in the prior literature on factors influencing customer retention and switching that customers stay with incumbent suppliers mainly due to the satisfaction felt about their services. In other words, customer satisfaction, as it positively affects customer loyalty, incites consumers to remain with their existing suppliers (Rust et al., 1993; Caruana, 2002). Customer satisfaction has proved a fundamental solution to prevent subscriber desertion, also in mobile communications services (Kim et al., 2004). However, there are cases in which customer behavior cannot be entirely explained by customer satisfaction alone. For example, some customers who are not satisfied with their current mobile service, nevertheless choose to stay with the existing carrier. Some customers switch to a new provider, even if they are satisfied with their incumbent provider. These are clearly cases in point suggesting that customer satisfaction is not a one-size-fits-all explanation for customer behavior.

Switching barrier is a concept introduced to explain cases of customers staying with incumbent suppliers, despite a low level of satisfaction. In recent years, this concept has been widely used in many empirical investigations of customer behavior in service industries, including mobile communications service. Concretely, switching barrier refers to the economic or non-economic cost switching entails, having the effect of dissuading customers from leaving the incumbent supplier, in spite of dissatisfaction; in other words, locking in customers.

There is, however, no equivalent concept to explain cases where satisfied customers, nevertheless, switch suppliers. Attempts to explain the phenomenon of customer desertion occurring in situations in which customers are satisfied have been surprisingly few. The phenomenon has been largely overlooked in favor, for instance, of lock-in among dissatisfied customers, a subject of much attention, both theoretical and empirical. Richard (1997) found that the ability to retain customers is crucial for service firms, as customer desertion has a tremendously negative impact on their company value. Hence, concepts like switching incentives may be needed to explain why satisfied customers change suppliers, as well as empirical investigations validating such concepts with actual data.

To understand the behavior of customers in a situation where they are heavily competed for, we need to move beyond the simple explanatory mechanism consisting solely of customer satisfaction and customer behavior as its two terms. What we need is finding a new construct which may serve as the intermediary between the two. In this study, we attempt to create a new understanding of how customer satisfaction influences customer behavior through two intermediary concepts: switching incentives and switching barriers. Our goal is to provide empirical evidence that customer satisfaction is not the only factor explaining customer behavior and that there are other variables that explain why in certain situations, customers' behavior is independent of the level of satisfaction. The market investigated in this study is the Korean mobile communications service market, where there is a high level of customer churn, as this maturing market is currently undergoing a transitional phase.

The rest of this paper is organized as follows: the Market Overview and Theoretical Background section offers an overview of competition among Korean mobile carriers and resulting customer behavior. The overview of the competition status is followed by a review of the existing literature on customer behavior, switching incentives and switching barriers, focusing more particularly on studies relating to mobile communications services. In the section on Research Model and Methodology, we describe the research model used in this study and methods, including the method of survey. The results of empirical analysis are presented also in this section, along with their interpretation. In the section on Directions for Future Research, we discuss ways in which future research can expand on this study, and the Conclusion section provides a summary of the results and their practical implications.

MARKET OVERVIEW AND THEORETICAL BACKGROUND

Overview of Korean Mobile Communications Service Market

The Korean mobile communications service market was opened to competition in the mid-1990s. Since then, the subscriber base for mobile communications services has been growing at a double-digit rate every year. The principal contributing factors to this galloping growth were the introduction of digital systems, massive build-up of service infrastructure, heavy investment in marketing by mobile carriers and the expansion of the lineup of mobile devices

This trend, however, as can be seen in the statistics provided in Table 1, has sharply slowed since the mid-2000s. Once the number of subscribers passed 70% of total population, the mobile sector started to exhibit the classic symptoms of a maturing market.

Yet, even as the market is fasting approaching the point of saturation, there has been an increasing volume of customer churn. The high level of customer churn was caused in large part by promotional marketing strategies adopted by Korean carriers, such as handsets offered at deep-discount prices or free replacement of handsets. The trend of carrier switching has been especially accelerating since the introduction in 2004 of number portability, a measure intended to dilute brand images of dominant suppliers and promote competition.

Since 2006, with the evolution of mobile services from 2G services to W-CDMA/ HSDPA (High Speed Downlink Packet Access)-based 3G services entering full swing, the competition is also intensifying between SK Telecom and KTF. At stake for the two W-CDMA service providers, clearly, is control over the 3G segment. KTF has currently managed to edge out SK Telecom which otherwise leads in the overall share of the mobile service market, in the 3G segment. At the end of 2007, the number of subscribers to KTF-provided 3G services amounted to 3,210,000, well above 2,500,000 subscribers using SK Telecom-provided 3G services.

The Korean mobile communications service market, therefore, exhibits both the characteristics of a maturing market and a transitional market. The competition currently taking place within it is centered more on winning over competitors' subscribers than developing new subscribers, and the focus of marketing efforts is on migrating existing 2G subscribers to 3G services.

Review of Literature on Customer Behavior in Mobile Communications Service Market

Mobile communications service customers' switching behavior is a topic which has received researchers' attention since only recently, namely starting from the early 2000s. Approaches to the subject in the existing literature are roughly three: one is investigating the relationship between customer satisfaction, customer loyalty and customer retention/switching, in the context of customer loyalty research. Studies demonstrating that satisfied customers, insofar as customer satisfaction is a determinant of customer loyalty and customer retention, are loyal to incumbent suppliers and remain with them belong to this first category. One case in point is the empirical analysis of customer behavior in the German mobile communications service market by Gerpott et al. (2001), who found that customer satisfaction, customer loyalty and customer retention mutually influenced each other. Kim et al. (2004) determined factors influencing customer loyalty in the Korean mobile communications service market and provided empirical evidence that customer satisfaction is a factor directly affecting the formation of customer loyalty. Lin and Wang (2006) demonstrated, using a structural equation modeling approach, that customer satisfaction was a determinant of customer loyalty in the Taiwanese mobile communications service market.

The second category of studies is concerned with the role of switching barriers and costs as intermediaries in the relationship between

Table 1. Subscriber trends in Korean mobile communications service (source: Websites of the Korean Ministry of Information and Communication and mobile carriers)

Year	2004	2005	2006	2007
Number of subscribers	3,659	3,834	4,020	4,350
Penetration rate	75.9%	79.4%	83.2%	89.8%
Number of new subscribers	299	176	186	330
Rate of subscriber growth	8.9%	4.8%	4.8%	8.2%
Number of number portability switchers	294	557	733	880
Percentage share of number portability switching in overall switching	8.0%	14.5%	18.2%	20.2%
Market share of SK Telecom	50.5%	50.9%	50.4%	50.5%
Market share of KTF	32.1%	32.1%	32.1%	31.5%
Market share of LG Telecom	16.4%	17.0%	17.5%	18.0%

(unit: ten thousand)

customer satisfaction and customer retention or loyalty. Works empirically testing the hypothesis that switching barriers or costs have a moderating effect on the influence of customer satisfaction on customer retention or customer loyalty fall into this category (Lee and Lee, 2005). Lee et al. (2001) empirically confirmed, in their study of French mobile communications service subscribers, that switching costs play the role of moderator in the relationship between customer satisfaction and customer loyalty. Lee and Kim (2003)'s empirical analysis of the relationship between customer satisfaction and customer loyalty in Korean mobile communications services also reported evidence indicating the existence of moderating effects of switching costs. Meanwhile, Kim et al. (2003), in their study of Korean mobile communications service customers, found that factors composing switching barriers had a moderating effect on the impact of customer satisfaction on customers' switching behavior. The investigation of US mobile communications service subscribers by Eshghi, Haughton and Topi (2007) reported that customer satisfaction, as a factor influencing customer loyalty and retention, had a moderating effect on switching.

The third and last type of research is focused on the influence of the policy environment on customer switching. A popularly researched topic in recent years has been the effect of the unification of carrier identification codes on subscribers' switching behavior. Kim et al. (2005) empirically established evidence that the introduction of number portability in Korea contributed to a sharp increase in switching intention among subscribers. Lee et al. (2006) predicted, using a conjoint analysis approach, that the introduction of number portability would increase the volume of subscribers switching carriers and intensify competition among carriers. Shin (2007) found evidence that the introduction of number portability in the US lowered switching costs and increased switching among American mobile subscribers.

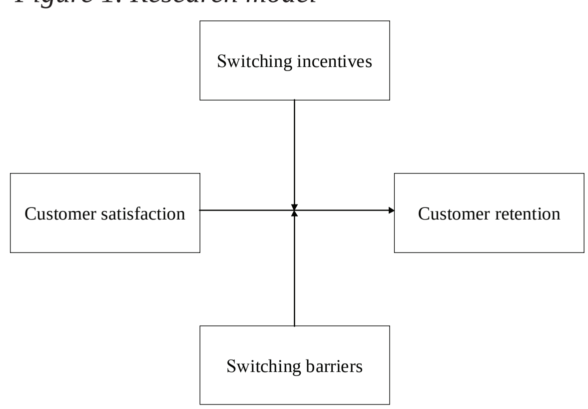
Customer retention and switching in mobile communications services being a topic researched since relatively recently, the existing literature is still rather scarce (Kim et al., 2004). Moreover, existing research is mostly concerned with determining influence factors for customer retention or the role of switching barriers in the relationship between customer satisfaction and retention. Few attempts have been made to comprehensively assess the relationship between customer satisfaction and retention. Studies rarely considered, for instance, the involvement of new switching factors such as handset subsidies and other promotional benefits, or service upgrades.

RESEARCH MODEL AND METHODOLOGY

Research Model

The goal of this study is to determine factors influencing customer retention in the Korean mobile communications service market, using theoretical constructs drawn from related literature. The study is more particularly concerned with casting light on the role played by switching incentives and barriers in the relationship between customer satisfaction and retention. To this end, we constructed the following research model:

Figure 1. Research model



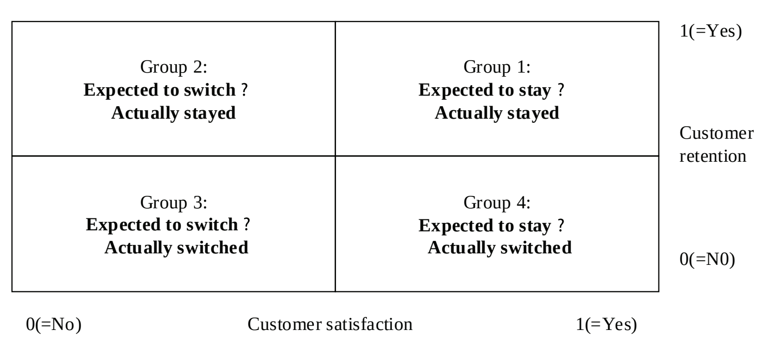
Methodology

The methodology used to empirically analyze the research model is as follows: mobile customers were segmented into subgroups, as shown in (Figure 2), according to customer satisfaction and carrier switching-related criteria. Members of group 1 in (Figure 2) are customers who are generally satisfied with their incumbent carriers and are remaining with them. Members of group 3 are customers who are dissatisfied with their incumbent carriers and who, therefore, switched to a new carrier. The behavior of both group 1 and group 3 subscribers conforms to the relationship between customer satisfaction and switching predicted in the existing literature. On the other hand, the behavior of group 2 and group 4 subscribers falls outside the expected pattern. The former are dissatisfied customers who, nevertheless, chose to stay with their incumbent carriers, and the latter are customers who switched carriers, even as they remain satisfied with the existing one. Whilst the behavior of group 2 customers may be explained through an existing construct, namely switching barriers, there has been thus far practically no attempt to explain the behavior of group 4 customers. In this study, we try to explain the relationship between customer satisfaction and switching using the construct of switching incentives. We compare factors influencing the behavior of customers belonging to each of the four groups in (Figure 2) to determine the role of switching incentives and barriers in the relationship between customer satisfaction and switching.

To empirically validate the research model, we selected a series of variables, as shown in the list in Table 2, by consulting experts.[1] For customer retention, the dependent variable, the respondents are asked to indicate whether they switched carriers since 2006. The variable, therefore, indicates whether or not there has been a change of carriers. Customer satisfaction, the independent variable, was measured with regard to the previous carrier in the case where a respondent changed his or her carrier since 2006. In the opposite case where there has been no change of carrier, customer satisfaction was measured and with regard to the current carrier. Switching incentives and barriers, the moderators, were also measured with regard to the previous or current carrier, depending on whether there has been a change of carrier. In other words, the respondents were asked to answer questions related to switching incentives and barriers, with regard to the previous carrier, if they switched carriers, and with regard to the current one, if they did not switch carriers.

Two sub-variables were created under switching incentives: handset replacement subsidies and service upgrade. In Korea, since the introduction of number portability, there has been a sharp increase in the volume of mobile customers switching carriers. This increase is in part due to aggressive marketing strategies used by carriers, offering handset subsidies or replacing handsets

Figure 2. Customer segmentation by the status of satisfaction and switching behavior



free of charge or at a symbolical cost. Another factor explaining this surge in customer churn is the evolution of mobile technology. The evolution from 2G services to 3G HSDPA services sparked off a spree of massive network investment, particularly by KTF, the latecomer of the Korean mobile market. Carriers also engaged vigorous marketing efforts to migrate existing subscribers toward the new services, through offers of service upgrade, among others. The attractiveness of alternatives is a variable whose influence on switching behavior has long since been recognized (Bendapudi and Berry, 1997; Kim et al., 2005). When an alternative supplier is attractive enough, even customers who are otherwise satisfied with the incumbent supplier may switch.

Sub-variables of switching barriers included the burden of having to change the number, burden of losing current benefits, burden of having to pay back discounts as a result of early termination of contract, burden of incurring additional costs, and the unattractiveness of alternatives. The burden of having to change the number, whilst it is partially resolved with the introduction of number portability, remains nevertheless. With the implementation of 010, the unified mobile prefix, 2G customers upgrading to 3G services offered by a third-party carrier must change their number to one starting with 010. The burden of losing current benefits refers to the burden felt by mobile subscribers about losing benefits from customer reward programs offered by their current carriers, such as their balance of reward points. The burden of incurring additional costs corresponds to the perceived burden from the one-time subscription charge paid by new subscribers (ranging from 30,000 won to 50,000 won).

All variables were measured through a discrete choice question whereby 0 is assigned to 'No' and 1 to 'Yes.' The questionnaire developed from these variables was then evaluated through a pilot test conducted on experts of relevant fields.

The population surveyed was mobile users residing in the Seoul metropolitan area, aged between 15 and 49. The survey was conducted by an outside research firm in 2007, through one-on-one interviews, and resulted in 500 valid responses. SPSS 12.0 was used for basic statistical analysis of the data. Demographic characteristics of the sample and their mobile usage characteristics were as follows: the sample was almost evenly divided by sex, with men accounting for 47% of total respondents and women 53%. By age group, the sample broke down to 10.0% aged 19

Table 2. Variables

Variable		Description
Customer satisfaction		D=1: Perceived satisfaction with the service provided by a mobile carrier
Customer retention		D=1: Continues to use the same carrier
Switching incentives	Subsidies toward handset replacement	D=1: Perceived influence of an alternative carrier's offer of handset subsidies on the customer's switching decision
	Service upgrade	D=1: Perceived influence of an alternative carrier's offer of service upgrade on the customer's switching decision
	Attractiveness of alternatives	D=1: Perceived influence of the attractiveness of an alternative carrier on the customer's switching decision
Switching barriers	Burden of having to change number	D=1: Perceived influence of the burden of having to change the number on the customer's switching decision
	Burden of losing current benefits	D=1: Perceived influence of the burden of losing benefits from the current carrier on the customer's switching decision
	Burden of incurring addition costs	D=1: Perceived influence of the burden of additional costs on the customer's switching decision

and less, 29.6% aged between 20 and 29, 32.0% aged between 30 and 39, and 28.6% aged 40 and older. By occupation, students represented 21.0% of total sample, administrative and clerical workers 22.0%, self-employed workers 22.0%, stay-at-home wives and mothers 12.0%, and others 23.0%. 52.4% of the respondents were SK Telecom subscribers, 32.4% KTF subscribers and 15.2% LG Telecom subscribers; this breakdown is virtually identical to the subscriber share distribution among the same three mobile carriers from the end of December 2007.

Results

Using the methods described earlier, we conducted a clustering analysis with regard to the variables of customer satisfaction and retention and obtained four subgroups. The basic demographic characteristics of the four subgroups are given in Table 3.

The analysis of variance (ANOVA, Duncan Test) revealed that differences existed between the four groups, at the level of demographic characteristics related to sex (male), occupation (college students and stay-at-home wives), as well as that of IT usage characteristics related to the experience of using mobile phone-based wireless internet and frequency of handset replacement. The percentages of men and college students were lower in Group 2 (expected to switch → actually switched) than in the rest of groups. Groups 2 members also had a lower level of experience using mobile phone-based wireless internet and a lower frequency of replacing their handsets, compared to the rest of groups.

As for the two groups of customers satisfied with their incumbent carriers, no other difference was observed between group 1 (customers who actually stayed) and group 4 (customers who actually switched), except that members of the latter cluster changed their handsets more frequently than those of the former cluster. On the other hand, group 3, dissatisfied customers who ended up switching their carriers, had a higher percentage of college students and more extensive experience using mobile phone-enabled wireless internet services. These users also changed their handsets more frequently than Group 2, the cluster of dissatisfied customers who are staying with their incumbent carriers. These patterns suggest that the frequency of replacement of mobile handsets has a certain influence on switching behavior, independently of whether a customer is satisfied with the service provided by their current carriers.

The results of the analysis of variance between the four groups of mobile users, conducted to determine the role of switching incentives and barriers in the relationship between customer satisfaction and switching, are given in Table 4.

Handset subsidies and the attractiveness of alternative carriers, two of the sub-variables under switching incentives, proved to have a greater influence on the switching behavior of group 3 and group 4, mobile users who actually switched their carriers, than that of group 1 and group 2.

Meanwhile, the burden of having to change the number and that of losing current benefits, two of the sub-variables under switching barriers, appeared to play a more significant role for group 1 and group 2, mobile users who are staying with their incumbent carriers, than for the rest of groups. In other words, the influence of switching incentives was greater among mobile customers who actually switched their carriers, whereas the impact of switching barriers was stronger among customers who are staying with their incumbent carriers.

If we compare the two groups of satisfied customers who, however, differ in their switching behavior, handset subsidies and the attractiveness of alternative carriers had a greater impact on members of group 4 (customers who actually switched carriers) than on members of group 1 (customers who are actually staying with the incumbent carrier). On the other hand, group-1 users responded more sensitively to switching

Table 3. Statistical characteristics of subgroups

		Unit	Overall	Group 1	Group 2	Group 3	Group 4	F-value	Duncan Test
N			500	212	38	178	72	-	-
Customer satisfaction		-	0.57	1.00	0.00	0.00	1.00	-	-
Customer retention		-	0.50	1.00	1.00	0.00	0.00	-	-
Demographic characteristics	Male	%	0.47	0.53	0.34	0.40	0.54		2<1**,4** 3<4*
	Age	Year	30.92	31.21	32.82	30.03	31.26	1.01	-
	Martial status	%	0.50	0.50	0.58	0.48	0.53	0.50	-
	Education	Year	13.67	13.67	13.47	13.62	13.90	0.56	-
	Monthly household income	KRW 10,000	383.89	388.49	372.37	381.57	382.15	0.29	-
	Apartment residents	%	0.56	0.52	0.63	0.58	0.61	0.97	-
	College students	%	0.13	0.11	0.05	0.15	0.17	1.28	2<3*,4**
	Administrative and clerical workers	%	0.22	0.24	0.26	0.19	0.22	0.53	-
	Blue collar and service workers	%	0.13	0.14	0.13	0.11	0.17	0.49	-
	Self-employed	%	0.22	0.20	0.18	0.25	0.21	0.68	-
	Stay-at-home wives/ mothers	%	0.12	0.15	0.18	0.11	0.06	1.86	4<2**
IT usage characteristics	Years using a mobile phone	Year	8.23	8.41	7.76	8.06	8.35	0.76	-
	Average monthly personal expenditure on mobile phone service	KRW 10,000	60.40	62.91	56.66	58.34	60.06	0.65	-
	Experience using mobile phone-based wireless internet	%	0.67	0.66	0.53	0.68	0.74	1.71	2<1*,3*,4**
	Frequency of mobile phone replacement		3.42	3.14	2.82	3.62	4.10		2,<3*,4*** 1<4**

**:p<0.1, **:p<0.05, ***:p<0.01*

barriers, such as the burden of having to change the number and that of losing current benefits.

The comparison of the two groups of dissatisfied customers reveals that switching incentives such as handset subsidies and the attractiveness of alternative carriers had a more significant impact among members of group 3 (customers who actually switched carriers) than among members of group 2 (customer who nevertheless stayed with the incumbent carrier). As for switching barriers such as the burden of having to change the number or that of losing current benefits, they appear to affect group-2 customers more significantly than group-3 customers.

To sum up, the role of switching incentives in the relationship between customer satisfaction and switching behavior consists in inciting satisfied customers to nevertheless switch carriers. Meanwhile, switching barriers influence the same relationship by dissuading dissatisfied customers from actually leaving the incumbent carrier.

Table 4. Results of analysis of variance between groups

		Overall	Group 1	Group 2	Group 3	Group 4	F-value	Duncan Test
Switching incentives	Subsidies toward handset replacement	0.42	0.16	0.18	0.64	0.79	67.94***	1,2<3***<4***
	Service upgrade	0.19	0.17	0.18	0.19	0.24	0.44	-
	Attractiveness of alternative carriers	0.33	0.24	0.21	0.39	0.54	9.51***	2,1<3**<4** 2,1<4***
Switching barriers	Burden of having to change number	0.18	0.25	0.34	0.11	0.10	7.79***	4,3<1**,2***
	Burden of losing current benefits	0.32	0.42	0.45	0.24	0.19	8.10***	4,3<1**,2***
	Burden of incurring additional costs	0.21	0.22	0.21	0.20	0.18	0.23	-

*:p<0.1, **:p<0.05, ***:p<0.01

DIRECTIONS FOR FUTURE RESEARCH

This study has been an attempt to empirically determine the role of switching incentives and barriers in the relationship between customer satisfaction and switching behavior, using Korean data. Future research can expand theoretical understanding of customer behavior and offer practical suggestions useful for customer strategies of mobile communications service providers by improving on the following areas:

First, to help mobile carriers design more precisely targeted customer strategies, customers can be divided into more varied segments. Future research should take into consideration, in particular, the fact that business strategies based on customer value are more effective than strategies aiming at simple quantitative expansions of subscriber base. Customers can, for instance, be classified into segments according to ARPU (Average Rate Per User) or the length of subscription period, so that the analysis of customer switching behavior may yield more targeted results, according to customer value.

Second, future research can include more varied variables of switching incentives and barriers than have been used in this study. Advertising, distribution, rate discounts and pricing structure, pre-marketing, brand effects and other factors related to the 4Ps of marketing may be considered switching incentives. The list of switching barriers may include procedures for termination of subscription, follow-up management, social influence and learning cost. Although not all of these variables can be considered at once, the list of switching incentive and barrier variables, nevertheless, needs to be expanded, by selecting new variables suited to the focus or topic of the study.

Finally, considering the many segments making up the communications service industry or comparing with other IT sectors could also be a promising avenue to exploit. Investigating the role of switching incentives and barriers by comparing the mobile communications service sector with IT service sectors where there is a fierce competition for subscribers such as broadband access service, mobile broadcast or IPTV, or the mobile handset market can yield important theoretical insights into the relationship between customer satisfaction and switching.

CONCLUSION

In the communications service sector, customer value doesn't just determine the short-term viability of a company, but is also a crucial condition for its continued competitiveness and growth. Hence, retaining existing subscribers and expanding the subscriber base by winning over competitors' customers are an indispensable part of all business strategies. Competition for subscribers is particularly fierce in the Korean mobile market, as this fast maturing market is also going through a transitional period. Major realignments are under way since the recent introduction of number portability, coinciding with the accelerating shift of mobile standards toward HSDPA-based services. Designing efficient customer strategies, more important than ever in such times of change, requires a clear understanding of what factors influence customer behavior. The consensus in the existing literature is that customer satisfaction is the principal precedent of switching behavior. However, customer satisfaction fails to explain customer behavior in certain cases. In other words, in some cases, satisfied customers do switch suppliers, and in some other cases, customers remain with their existing suppliers, despite the dissatisfaction with the service. In this study, we attempted to explain these types of customer behavior that customer satisfaction alone cannot account for. To achieve this goal, we used two additional constructs: switching barriers, an existing concept, and switching incentives, a concept that has received so far much less attention than the latter. The principal findings of this study are as follows:

First, switching incentives had a moderating effect on the relationship between customer satisfaction and switching; in other words, a satisfied mobile customer's decision to switch carriers was influenced by switching incentives. Our results suggested that in the Korean mobile communications service market, satisfied customers do leave their incumbent carrier for an alternative carrier, if the offer by the alternative carrier is attractive enough (ex. attractive handset subsidies). However, even as 3G services are now becoming the norm of the mobile market, features like the possibility of video conference or faster wireless internet access fail to play the role of a switching incentive. The recent sharp increase in number of HSDPA subscribers in Korea, however, has not resulted in any commensurate increase in number of subscribers actually using video conference or wireless internet, at least according to data from early 2008 (Kim and Park, 2008). Hence, to effectively win over competitors' customers who are satisfied with the service of their current carrier, a carrier should either make its services more convenient or offer the latest models of handset at no additional charge or at a deep-discount price. Offering more competitive rates than competitors or a superior quality of service should also incite mobile users to switch.

Second, switching barriers appeared to have a moderating effect on dissatisfied customers' decision to switch to a new carrier or remain with the incumbent carrier. The implementation of number portability has dramatically reduced the effectiveness of phone numbers as a switching barrier in the Korean mobile market. However, mobile users are still required to change to a number starting with the prefix 010, when they migrate from a 2G service to a 3G service. Due to this requirement, mobile phone numbers do remain a switching barrier, in spite of all. Our results also indicate that benefits from customer reward programs, handset discounts or rate discounts received in exchange for buying a long-term contract also contributed to locking in customers. The implication of this result is that a mobile operator should consider making its customer reward programs more attractive than those of its competitors, in order to win over the latter's dissatisfied customers.

Finally, there were some differences in terms both of demographic and IT usage characteristics between customer segments that were grouped according to the status of satisfaction and switching behavior. These differences between groups were notably related to sex, occupation, experience using mobile phone-enabled wireless internet and the frequency of handset replacement. Accordingly, it may be useful to integrate a CRM (Customer Relationship Marketing) approach, taking demographic and IT usage characteristics of customers into close consideration in devising customer strategies.

REFERENCES

Abdolreza E., Dominique H., & Heikki T. (2007). Determinants of customer loyalty in the wireless telecommunications industry. *Telecommunications Policy, 31*(2), 93-106.

Bendapudi, N., & Berry, L. L. (1997). Customers' motivations for maintaining relationships with service providers. *Journal of Retailing, 73*(1), 15-37.

Gerpott, T., Rams, W., & Schindler., A. (2001). Customer retention, loyalty, and satisfaction in the german mobile cellular telecommunications market. *Telecommunication Policy, 25*(4), 249-269.

Lee, H. S., & Kim, S. Y. (2003). A study on the moderate effect of switching cost and the customer satisfaction affecting customer loyalty. *Telecommunications Review, 13*(6), 1027-1039.

Lee J., Lee, J., & Freick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link; Mobile phone service in france. *Journal of services Marketing, 15*(1), 35-48.

Lee, J. S, Kim, Y. B., Lee, J. D., & Park, Y. R. (2006) Estimating the Switching Cost and its Implications for Market Structure: Case of the Korean Mobile Communications Market and Number Portability. *The Korean Journal of Industrial Organization, 12*(2), 33-58.

Lin, H. H., & Wang, Y. S. (2006). An examination of the determinants of customer loyalty in mobile commerce contexts. *Information & Management, 43*(3), 271-282.

Jones, M. A., Mothersbaugh, D. L., & Betty, S. E. (2000). Switching barriers and repurchase intentions in services. *Journal of Retailing, 76*(2), 259-272.

Kim, M. K., Lee, M. H., & Park, J. H. (2005) Impacts of the introduction of mobile number portability on customer switching intention. *Telecommunications Review, 15*(4), 667-679.

Kim, M. K., & Park, J. H. (2008). Characteristics of subscriber and usage for WiBro and HSDPA at the introductory market in Korea. *Korean Radio Promotion Association,* Arpil-March.

Kim, M. K., Park, M. C., & Jeong, D. H. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in the Korean mobile telecommunication services. *Telecommunications Policy, 28*(2), 145-159.

Kim, M. K., Park, M. C., Jeong, D. H., & Park, J. H. (2003). An empirical study on the adjustment Effect of switching barriers on customer retention in mobile telecommunications services. *The Journal of MIS Research, 13*(3), 107-130.

Richard. S. (1997). Combating the churn phenomenon. *Telecommunications.*

Rust, R. T., & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing, 69*, 193-215.

Shin, D. H. (2007). A study of mobile number portability effects in the United States. *Telematics and Informatics, 24*(1), 1-14.

Yi, Y. H., & Lee, C. L. (2005) Antecedents and consequences of switching costs: The moderating role of service subscription Types. *Journal of Korean Marketing Association, 20*(3), 1-28.

WEB SITES

http://www.mic.go.kr

http://www.sktelecom.com

http://www.ktf.com

http://www.lgtelecom.com

KEY TERMS

Customer Churn: An opposite concept to customer retention referring to cases in which a carrier's subscribers desert it for alternative carriers. Customer dissatisfaction is believed to be the antecedent of customer churn. Customer churn and customer switching are interchangeable terms.

Customer Retention: A subscriber is considered to have been retained by his/her current carrier, when he/she has not left it for an alternative carrier. According to the literature, customer satisfaction is an antecedent of customer retention.

Customer Satisfaction: A customer's overall perception of service quality, price, customer care and others indicating the level of satisfaction with a carrier.

HSDPA (High Speed Downlink Packet Access): A WCDMA-type, 3-G mobile access technology, providing faster speeds of data download than 2-G technologies like CDMA and GSM. Services offered through HSDPA include video call, video conference and wireless broadband internet access.

Mobile Number Portability: A policy allowing mobile users to retain their current numbers, even after they switch carriers. The policy is intended to curb brand effects from number and eliminate a major switching barrier for mobile subscribers.

Switching Barrier: A factor which hinders a mobile subscriber from switching carriers. Switching barriers can dissuade dissatisfied customers from leaving their incumbent carriers.

Switching Incentive: A factor that motivates a mobile subscriber to switch his or her carrier. Switching incentives can incite even those customers who are satisfied with their current carriers to switch to a new carrier.

ENDNOTE

[1] Five experts working for a public research institution specialized in IT policy or business since three years or longer, in a position of senior researcher or higher, were consulted.

Chapter XXIII
Telecommunication Customer Demand Management

Jiayin Qi
Beijing University of Posts and Telecommunications, China

Yajing Si
Beijing Wuzi University, China

Jing Tan
Beijing University of Posts and Telecommunications, China

Yangming Zhang
Beijing University of Posts and Telecommunications, China

ABSTRACT

In the highly competitive environment, the real-time and dynamic customer demand analysis technology is required by the enterprise information systems in order to respond to customer demand efficiently and automatically. Based on a customer value hierarchy model, this chapter proposes a customer demand analysis method and proposes ways to capture customer demand knowledge. Then, we present a novel product recommendation approach, which involves the customer value hierarchy model into traditional recommender systems. Through the above steps, the telecom operators can get their customer demand and respond to their demand automatically.

INTRODUCTION

Telecommunication industry is a special industry. The products and services in this industry are immaterial, complicated and technical. Nowadays, telecommunication operators are facing more challenges. Globalization and technology innovation present radical challenges to telecom operators. The telecommunication operators must be more and more competitive in order to survive in the future telecommunication industry (Zhang et al., 2004; Wang & Archer, 2007).

Today's telecommunication market encourages more competitions. It offers more choices for customers, such as lower price, and the improved service quality. As the previous monopoly situation no longer exists, new entrants come into the market. In an emerging economy, state-owned operators are fully or partially privatized in order to survive better (Stienstra et al., 2004).

Globalization also promotes the domestic competition. Global telecommunication market provides opportunities to some operators because of the economies of scale in telecommunication networks, such as BT and Vodafone. It also brings radical domestic competition since more new operators enter the market.

Internet technology causes an extraordinary growth of the Internet and IP services and applications. Customers are increasingly free to choose different service components from different vendors and assemble their own solution (Li & Walley, 2004). 3G (the third generation telecommunication technologies) and mobile Internet accelerate the production of new services (Pikkarainen, 2001; Keryer, 2001; Meier et al., 2003).

Industry deregulation, globalization, and IP make the telecommunication industry be full of intensified competition. The telecommunication market involves a shift from a stable market to an increasingly user-driven market place. The success of a telecom operator will entirely depend on the operator's ability to create services and applications that are embraced by the users.

The effective management of basic telecommunication infrastructures, full range of software platforms, increasing varieties of services, and a large scale customer base are very important for any operator who wants to win the market. The telecom operators must have the ability to rapidly develop, deploy, and manage services to meet customers' dynamic requirements any time and any where.

The innovation of the telecommunication products mostly depends on the mutual actions between telecom operators and customers. Therefore, Telecom operators should understand **customer demands** to know what customers want, and then provide them the products or services they need as well as possible. This is an essential way to improve customer satisfaction, upgrade customer value, acquire loyal customers, enlarge telecommunication market share, and maximize the operator's revenue, and shareholders' value.

Telecommunication customer demands can be managed effectively by analyzing customer psychology and behavior. By analyzing customer psychology, operators may radically understand the consuming motivation and purpose of the customer. However, the customer psychology is recessive, and is difficult to be obtained. In addition, customer behavior also can be used to analyze customer demands. Most importantly, it is obvious and can be obtained easily. However, the analysis based on the customer behavior cannot always reach the deep layer of customer demands. Therefore, the real-time and dynamic **customer demand** analysis technology is required for capturing customer demand knowledge.

Influenced by the consuming aim, consumption environment and individual preference, different customer groups have different customer demands (Sharp, 1997; Boulton, 2000). Previous studies mainly focus on these subjects: firstly, predicting customer preferences and repeat-purchase patterns through consume history analysis (Simpson, 2001); secondly, analyzing the antecedents and consequences of consumer behavior and **customer loyalty** (Srinivasan, 2002; Inoue, 2003); thirdly, classifying customers using clustering analysis (Wan, 2005). It's a bit excessively simple with such a low intelligence level and large manual work. Woodruff, Butz and Goodstein proposed the CVD (Customer Value Determination) model and built the correlative relations among the **customer demand** attribute layer, the consequence layer and the objective layer (Burs, 1991). However this research did not present technical tools to implement the CVD knowledge capture.

On the basis of the customer demand analysis, the design of **personalized recommendation** systems is another important work in the telecommunication customer management. The **personalized recommendation** system is based on customer's demand, interest or behavior mode, and actively commends some individual products, services or information to customers who are interested in those commodities (Chai, 2006). With the development and popularization of intelligent mobile terminal, and with the research and promotion of ubiquitous computing technology (U-computing), operators gradually restructure from network provider to information provider. Under the environment of ubiquitous network (U-network), the telecommunication service **personalized recommendation** will enhance telecom operator's ability of customer management.

Current common approaches for **personalized recommendation** systems are the content-based approach and collaborative filtering approach (Boulton, 2000; Inoue, 2003). Content-based systems provide recommendations by matching customer interests with product attributes. On the other hand, collaborative systems utilize the overlap of preference ratings among customers for product recommendation. Existing content-based systems were confined owing to the contingency upon the customer interests, resulting in products being recommended homogeneously. Collaborative systems advance content-based systems in their capabilities of recommending heterogeneous products (e.g. beer and diaper example). Both approaches are based on data-driven analysis and on the assumption that similar customers make similar choices. Research on customer purchase intention, however, is excluded from these approaches. Woodruff defines the conception of **customer value** from the perspective of customers (Wan, 2005). **Customer value** is a perception of what a customer wants to accomplish with the help of products, in order to reach a desired goal. He argues that **customer value** describes the nature of the relationship between the customers and the products, while customer satisfaction represents a customer's reaction to the value received from the products. By constructing the customer value hierarchy for target customers, products suggested by a recommender system will please customer and be accepted in all probability, since they reach customers' desired goals.

Telecommunication customer demand management is an important part in the practices of telecommunication customer management. It helps telecom operators to develop the right products or services for customers. It will become enterprises' core strength and resources that competitors can not imitate.

This chapter is structured as follows: In Section 2, a **customer value hierarchy** model proposed by Woodruff is introduced. This model is used to analyze customer demand. In Section 3, on the basis of the Woodruff model and data from mobile communication customers, an intelligent method is presented to identify telecommunication customer demand. In Section 4, a telecommunication customer personalized recommendation system is developed based on the telecommunication customer demand identification. At the end of this chapter, the problems of dynamic detecting of customer demand and developing new telecommunication products or services according to the customer demand will be discussed with the direction of the future telecommunication customer demand research.

BACKGROUND

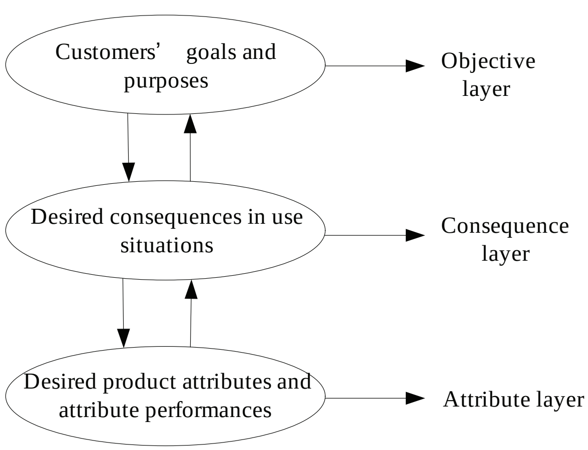
Concept of Customer Value

In this chapter, the concept of the **customer value** is defined from the perspective of a customer. Woodruff (1997) argues that **customer value** is a customer's perceived preference for and evaluation of those product attributes, product utilities, and consuming consequences arising from using the product to facilitate achieving the customer's goals or purposes in some situations.

This definition incorporates both desired and received value and emphasizes that value stems from customers' perceptions, preferences, and evaluations. It also links together products with use situations and related consequences experienced by goal-oriented customers. This definition depicts a means-end type of model, which was originally intended to describe how customers categorize information about products in memory (Burs, 1991). Figure 1 illustrates this model which can be adapted to capture the essence of **customer demand**.

The **customer value hierarchy** suggests that customers conceive of desired value in a means-end way. Starting at the bottom of the hierarchy, customers learn to think about products as bundles of specific attributes and attribute performances. When purchasing and using a product, they form desires or preferences for certain attributes based on their abilities to facilitate achieving desired consequence experiences, reflected in use value and possession value, in the next level up in the hierarchy. Looking down the hierarchy from the top, customers use goals and purposes to

Figure 1. Customer value hierarchy model



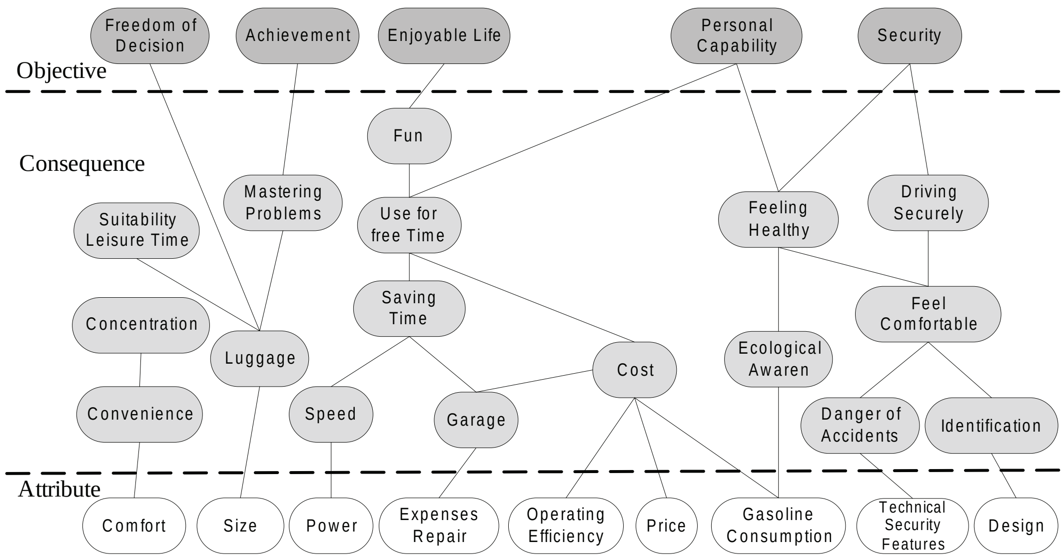
attach importance to consequences. Similarly, important consequences guide customers when attaching importance to attributes and attribute performances.

Determination of Customer Value Hierarchy Model by CVD method

In order to uncover the attribute-consequence-goal chain, a depth interview technique called "laddering" was developed (Reynolds, 1988). Subsequently, Walker and Olson developed a paper-and-pencil version (Walker & Olson, 1991). Laddering refers to a 2-stage process. (1) Elicitation of salient criteria used to discriminate between products. This is feasibly done by means of direct questioning. In this stage, concrete attributes (e.g. price) and abstract attributes (e.g. efficient) are identified. (2) The salient attributes (concrete or abstract) that are elicited form the starting point for the laddering probes which reveal the entire means-end structure. This is completed by continuous probing with some form of the question "Why is that important to you?" The response at each level is used as the basis for further questioning. This repetitive questioning is meant "to force the subject up the ladder of abstraction" until the goal level is reached.

The complete chain of product attributes, desired consequences and customer goal elicited by one individual constitutes a "ladder". The next step in the procedure involves a shift from the ladders produced by many individuals to the aggregate cognitive structure of a group of people. First, each idiosyncratic laddering response is assigned to the appropriate category of meaning at a particular level of abstraction (attribute, consequence or goal). Next, an Implication Matrix is computed by specifying the number of times two adjacent levels in the ladder which were connected by some respondents. This matrix allows the construction of a **customer value hierarchy** which graphically represents the aggregate meaning structure of the group. Figure 2 illustrates a customer value

Figure 2. Customer value hierarchy regarding automobile (Mark, 2004)



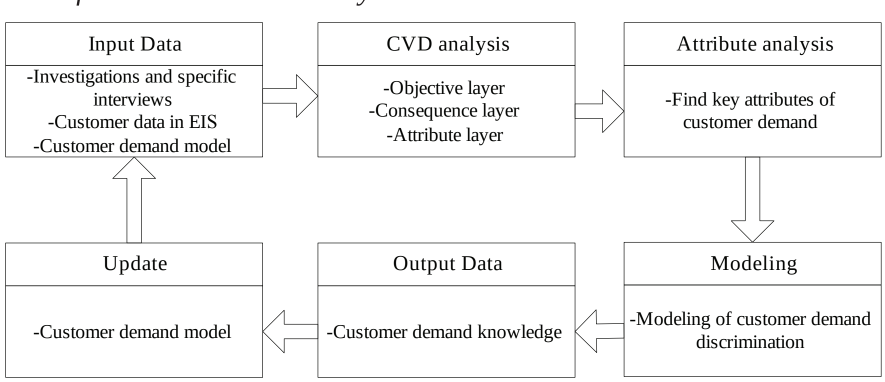
hierarchy regarding automobile as a result of a study conducted in Germany (Mark, 2004).

CUSTOMER DEMAND DISCRIMINATION WITH ARTIFICIAL INTELLIGENCE

As we have said, the CVD method proposed by Woodruff, Butz and Goodstein is an effective way to depict customer demand from three layers—attribute layer, consequence layer and objective layer. But the laddering interview method is not an efficient way to get each customer's demand. Some **artificial intelligence** techniques are required to deal with the matter and to improve the efficiency automatically (Qi et al, 2007). Combined the laddering interview with some **artificial intelligence** techniques, an analysis process for capturing customer demand is presented in Figure 3.

Input data. The input data may include the result of specific investigations and interviews for some customer samples, customer purchasing record and other related data in the enterprise'

Figure 3. Process of customer demand analysis



information systems and foregone **customer demand** model.

CVD analysis. According to Woodruff's CVD theory, which suggested that customer demand hierarchy contains objective layer, consequence layer and attribute layer, we define telecommunication **customer value hierarchy** and built the correlative relationships among layers.

Attribute analysis. This step is to find the significant attributes of customer value layers, a group of attribute layer variables which influence the **customer demand** objectives. This step is mainly to decrease the data dimension in customer demand analysis, and that is solved by the fuzzy cluster analysis in our research.

Modeling. In this step, the customer demand analysis model is built by adopting a neural network method.

Output. A well-formed model could judge the classification of customer demand objective layer perfectly from their choices on demand attribute layer.

The subsequent parts of this section will explore the process in details. For we have applied the above analysis process in several mobile operators in China, we would prefer explaining each steps in the process with examples from our practice in mobile telecommunication, although the method can be commonly used in the entire telecommunication industry to understand customer demand by **artificial intelligence** technique.

Data Preparation and the CVD Analysis Definition of Customer Value Layer in Telecommunication Industry

According to the **customer value hierarchy** model, we define that the telecommunication customer value hierarchy consists of the customer demand objective layer, the consequence layer and the attribute layer.

Objective layer. The objective layer means the ultimate motivations of customers engaging in telecommunication services. Customers may have multiple motivations in the objective layer.

Consequence layer. The consequence layer means the customer experience of telecommunication services.

Attribute layer. The attribute layer means the usage of telecommunication services/products.

Data Collection: An Example in Mobile Telecommunication Industry

This chapter takes the mobile individual customers of X city in China as the sample set of the investigation. We distributed 150 questionnaires. 120 effective questionnaires were returned. The rate of retrieving efficiency is 81.3%.

The questionnaire contains four parts: (1) questions about the customer demand objectives and the importance of them. A five-point scale was adopted to fill out the questions: one means the least important and five means the most important; (2) questions about the customer's expectation of consumption experience according to their demand objectives; (3) questions about which services/products serve the customer's demand objectives; (4) questions about whether the customers have used the products/services of the attribute layer in order to achieve their objectives. The questionnaire enumerates products/services of the attribute layer corresponding to a given objective of the objective layer. 1 means the customer has used the products/services and 0 means the customer hasn't used the products/services.

Mobile Customer Value Hierarchy: An Example in Mobile Telecommunication Industry

Based on the questionnaire survey and customer laddering interview that made in X city, we constructed the mobile **customer value hierarchy**

Table 1. Mobile customer value hierarchy in X city

Objective layer	Consequence layer	Attribute layer	
Communicative Object (a_{26})	Convenient communication,	short massage service call waiting call diversion little secretary voice mail box	a_1 a_2 a_3 a_4 a_5
Business Object (a_{27})	High quality, knight service, efficiency	U-net Routine service Ticket booking Uni-colour E E-bank Stock exchange Mobile purchase	a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12}
Recreational Object (a_{28})	Fashion, novelty, personality, fun	Color ring back tone mobile ring mobile picture E-game chat mobile movie	a_{13} a_{14} a_{15} a_{16} a_{17} a_{18}
Informational object (a_{29})	knowledge, timely information	News service Weather info Travel info Finance info Physical news Entertainment info U-map	a_{19} a_{20} a_{21} a_{22} a_{23} a_{24} a_{25}

with the CVD method which is presented in Section 2.2. The factors of the objective layer and attribute layer are defined as the variable a_i (i = 1, 2,...,29). See Table 1.

Significant Attributes of Customer Value Hierarchy

The significant attributes of **customer value hierarchy** mean the key attribute variables of the attribute layer which distinctly correlate with the objective layer. This step is mainly to decrease the data dimension in customer demand analysis.

The Principles of Significant Attributes Analysis

According to the rough set theory, data of the customer value objective layer and attribute layer can be defined as S = (U, A, V, f). Here,U = {u_1, u_2,...,u_m}where m is the total number of customers and U is the set of customers. A = {a_1, a_2,...,a_m}: The set of variables of the objective layer and the attribute layer. A = C ∪ D, where C is the characteristics set of the attribute layer, and D is the characteristics set of the objective layer. V is the set of the customer attribute parameters. The value of $f(u_j, a_i)$ indicates the value of u_j about a_i. The significant attributes are identified by fuzzy cluster. The process is as follows:

Step 1. Partition set A into C and D. Consider the numerical characteristics of attribute a_i in the attribute set C, and represent attribute a_i as a_{ij}(j = 1, 2,...,k). Here k is the number of incoordinate values of attribute a_i.

$$r_{ij} = \sum_{k=1}^{m}(a_{ik}a_{jk})^2 \Big/ \sqrt{(\sum_{i=1}^{m} a_{ik}^2)(\sum_{k=1}^{m} a_{jk}^2)} \quad (1)$$

Table 2. Significant attributes of the attribute layer

objective layer	λ	attributes cluster	Significant attributes
Communicative Object	0.908	{a1}{a2,a3,a4}{a5}	a1,a2,a3,a4
Business Object	0.999	{a6,a9},{a7},{a8,a10,a11,a12}	a6,a7,a9
Recreational Object	0.997	{a13,a14}{a15},{a16}{a17,a18}	a13,a14,a15,a16
Informational Object	0.95	{a19},{a20},{a21,a22,a23,a24,a25}	a19,a20

Step 2: Calculate the fuzzy similarity matrix *R*.

Step 3: Calculate the fuzzy transitive closure $t(R)$ of the fuzzy correlation matrix *R*. Use the cluster method to analyze $t(R)$ with intercept λ and find out the significant attributes set.

The Process of Data Analysis: An Example in Mobile Telecommunication Industry

In order to be less costly and easily applied, 50 questionnaires are chosen as the analysis samples. From total 25 products/services, 13 products/services were found that have distinct correlation with the customer demand objective. The results are presented in Table 2.

Therefore, the set of significant attributes C_c can be expressed as:

$$C_c = \{a_1, a_2, a_3, a_4, a_6, a_7, a_9, a_{13}, a_{14}, a_{15}, a_{16}, a_{19}, a_{20}\}$$

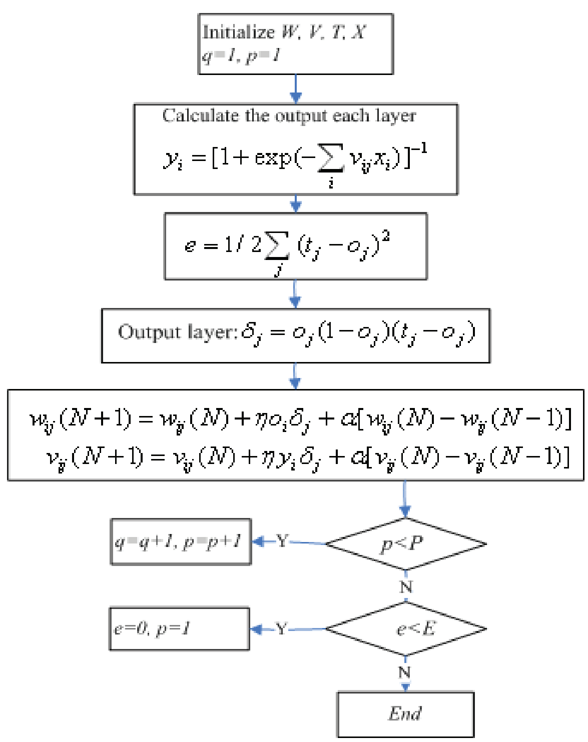
Telecommunication Customer Demand Discrimination Modeling

The customer demand hierarchy is applied into customer demand analysis modeling to capture **customer demand** object by analyzing customer demand attribute. We achieve this well by adopting a BP-neural network method.

The Principle of Modeling of Customer Demand Discrimination

We build the customer demand discrimination model by making the customer demand objective layer $U = [u_{1k}, u_{2k},...,u_{nk}](k \in D)$ as the output of the model and making significant attributes of the attribute layer $U = [u_{1j}, u_{2j},...,u_{nj}](j \in C_c)$ as the input of the model. The method of a **neural network** is shown in Figure 4.

Figure 4. The model of a BP-neutral network for customer demand discrimination



Step1: Initialization. Make the customer demand objective layer $U = [u_{1k}, u_{2k},...,u_{nk}](k \in D)$ as the desired output of the model, which is redefined as T. Make the set of customer demand significant attribute $U = [u_{1j}, u_{2j},...,u_{nj}](j \in C_c)$ as the input training sample set of the model, which is redefined as X. p is the number of the training sample.

Step 2: Calculate the actual output. Calculate the output of each layer with the function

$$y_i = [1 + \exp(-\sum_i v_{ij} x_i)]^{-1}, \quad , \tag{2}$$

$$o_i = [1 + \exp(-\sum_i w_{ij} y_i)]^{-1} \tag{3}$$

Where $X = (x_1, x_2,...,x_n)$ represents the input training sample and $T = (t_1, t_2,...,t_n)$ represents the desired output. The final actual output is written as $O = (o_1, o_2,...,o_n)$.

Step 3: Adjust the weight. Adjusting the weight with recursion method from output layer to hidden layer as the following formula:

$$w_{ij}(N+1) = w_{ij}(N) + \eta o_i \delta_j + \alpha[w_{ij}(N) - w_{ij}(N-1)] \tag{4}$$

$$v_{ij}(N+1) = v_{ij}(N) + \eta y_i \delta_j + \alpha[v_{ij}(N) - v_{ij}(N-1)] \tag{5}$$

Where o_i is the output of anterior layer, $0 \prec \alpha \prec 1$.

If j is in output layer, then

$$\delta_j = o_j(1 - o_j)(t_j - o_j) \tag{6}$$

If j is in hidden layer, then

$$\delta_j = o_j(1 - o_j)\sum_k \delta_k w_{jk} \tag{7}$$

Step 4: Consider the error e between the actual output vector $(o_1, o_2,...,o_n)$ and the desired output vector of sample, $e = \frac{1}{2}\sum_j (t_j - o_j)^2$, repeat step3, till the error e is acceptable.

Customer Demand Knowledge Acquisition: an Example in Mobile Telecommunication Industry

The model in this chapter can help the telecom operators to carry out more efficient "one to one" customer service strategy. Taking the mobile operator we cooperated as an example, the well-formed model could capture the **customer demand** knowledge from these aspects:

Judge the classification of customer demand objectives dynamically from their demand attributes. 3 customers are randomly selected from the samples and the comparison between the analytical conclusions and the actual demands is presented in Table 2. For getting obvious conclusion, we assume 3 as the dividing line. It can be indicated that the model can discriminate the **customer demand** with a high accurate percentage.

Get the sort order of customer demand objects. Table 2 presents the sort order of customer demand objects that are discriminated by the model, and that is accordant to customer's actual object's sort order.

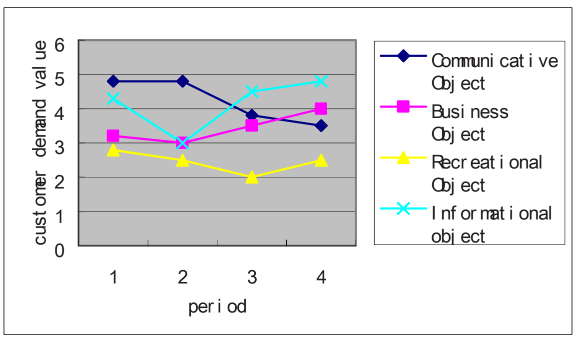
The model can indicate a measurable conclusion based on customer motivation that has the advantage of high stability and reliability.

Furthermore, by adopting the model to analyze multi-period customer data, the mobile operators can find the track of customer's demand transformation. An example of the customer's demand transformation track that was found by this method is presented in Figure 5.

Table 3. Comparison between analytical conclusions and actual demands

Customer		a_{26}	a_{27}	a_{28}	a_{29}	customer demand object
1	Actual demand	5.0	3.0	3.0	5.0	Communicative, Business , recreational, informational object
	Analytical conclusion	4.9	2.5	3.0	3.9	Communicative, recreational, informational object
2	Actual demand	5.0	3.0	2.0	4.0	Communication object Business object, Information object
	Analytical conclusion	4.9	3	2.9	3.8	Communication object Business object, Information object
3	Actual demand	5.0	2.0	2.0	3.0	Communication object, Information object
	Analytical conclusion	4.9	3.1	2.6	3.8	Communication object Business object Information object

Figure 5. A customer demand transformation track



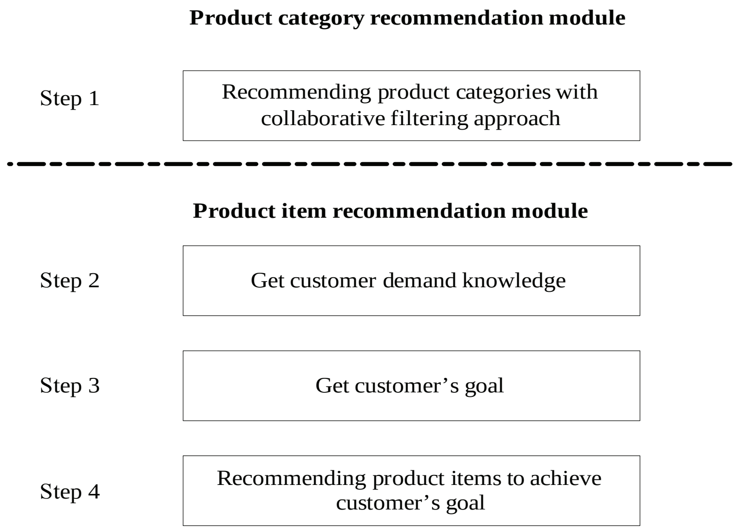
Methodology Comparison: An Experiment in Mobile Telecommunication Industry

For analyzing the influence of reducing the inconsequential attribute variables, we make all the attributes of the attribute layer as the input of the **neutral network** model and accomplish the process. The well trained model of customer demand discrimination of all attributes is used to analyze the 61 samples of the contrastive sample

Table 4. Methodologies compare

Customer demand discrimination model		discriminate with significant attributes	discriminate with all attributes
accuracy (%)		80.517	79.71
Communicative object (a_{26})	mean-root-square error	0.543	0.549
Business object (a_{27})	mean-root-square error	1.361	1.320
Recreational object (a_{28})	mean-root-square error	1.155	1.163
Informational object (a_{29})	mean-root-square error	1.220	1.182

Figure 6. Recommendation approach based on customer demand discrimination



set. Comparing with the discrimination model trained by significant attribute, the accuracy of demand analysis for the four objects is contrasted in Table 4. The customer demand discrimination model trained by significant attributes is not worse than the model trained by all attributes at all.

RECOMMENDATION APPROACH BASED ON CUSTOMER DEMAND DISCRIMINATION

The proposed recommendation approach consists of two modules, recommending product categories and providing specified product items. This can be called **Personalized Recommendation**. Take the beer/diaper story for example; once a customer buys beers, the product category recommendation module predicts he will buy diapers at the same time, while the product item recommendation model advises the brand of diaper he might choose. These two modules consist of four steps as shown in Figure 6.

Recommending Product Category

Collaborative filtering (CF) is the most successful recommender system technology to date, and is used in many of the most successful recommender systems on the Web. CF is involved in our proposed recommendation approach to recommend product category to a target customer based on the opinions of other customers. Assumed that historical purchasing transactions of n customers on m product categories are collected, the data is represented as a $m \times n$ customer-product category matrix $R = (r_{i,j})_{n \times m}$, such that $r_{i,j}$ is one if the ith customer has purchased the jth product category, and zero, otherwise.

The next step is that of computing the similarity between customers based on the above matrix as it is used to form a proximity-based neighborhood between a target customer and a number of like-minded customers. The proximity between two customer a and b can be measured by computing the Pearson correlation, which is given by

$$sim(a,b) = corr_{ab} = \frac{\sum_i (r_{ai} - \overline{r}_a)(r_{bi} - \overline{r}_b)}{\sqrt{\sum_i (r_{ai} - \overline{r}_a)^2 \sum_i (r_{bi} - \overline{r}_b)^2}} \tag{8}$$

The goal of this measure is to find, for each customer *u*, an ordered list of *l* customers $\Omega = \{N_1, N_2,...,N_l\}$ such that $u \notin N$ and $sim(u, N_1)$ is maximum, $sim(u, N_2)$ is the next maximum and so on.

Finally, the recommender looks into the neighborhood Ω of a customer and for each neighbor scans through his/her purchase data and performs a frequency count of the product categories. After all neighbors are accounted for, the recommender sorts the product categories according to their frequency count and simply returns the *N* most frequent product categories that have not yet been purchased by the active customer.

Get Customer Demand Knowledge

Using the above method in Section 3, we can get any customer's demand knowledge.

Discovering Customer's Goal

Customer buying behavior is often depicted as purposeful and goal-oriented. Recommender systems intend to provide products satisfying each customer's goal. Therefore, the core of our approach is to uncover the relationship between product attributes and customer's goals using the customer value hierarchy. From the **customer demand** knowledge analyzed in Section 3, it is easy to discover a customer's goal.

Recommending Product Item

Recommendation system decides which brand and style of product in the same category (called product item in the article) should win the favor of a customer whose purchase goal has been identified in the prior step. Suppose performances of product attributes have been scored in advance by product experts. After all product satisfaction rates to a customer's identified goal are accounted for, our system sorts the products according to their satisfaction rates and derives the top-N recommendations.

FUTURE TRENDS

The traditional telecommunication voice services meet customers' requirement of security and communication. With the development and promotion of more new telecommunication value-added services, it will meet customer's spiritual requirements, such as information capture and entertainment, etc. How to obtain these spiritual requirements will become an important part of telecommunication operation and future research in this field. This chapter is an attempt in this research field. Nowadays, change is the permanent topic. The **customer demand** acquisition should be captured in the dynamic way. With the rapid increase of the number of mobile phone users, mobile phone will become an important channel to catch the data of **customer demand**. The continual development of the U-network and U-computing will offer a new way to catch the data of mobile **customer demand**. On the foundation of the dynamic acquisition of customer demand knowledge, the **personalized recommendation** will become a hot topic in this field.

CONCLUSION

The telecommunication customer demand management is an important part of telecommunication customer relationship management. The effective capturing of the data of customer demand is the foundation of customer demand management, and the purpose of customer demand management is to respond to the customer demand timely and accurately. In this chapter, we propose a semi-structured method of getting customer demand knowledge. The first step: according to the CVD method in the Woodruff model, get the telecommunication customer demand hierarchy model through an in-depth interview; the second step: get the significant attribute variables with the Fuzzy cluster method; the third step: Use these significant attribute variables as input into the BP-**neutral network** method to obtain the customer demand knowledge, and finally modeling the customer demand knowledge. When one telecom operator applies the customer demand knowledge model, it can first use the CVD to get the demand knowledge from the sample customers, then use these demand knowledge as training to get the demand knowledge of other customers by machine learning technique. We also discuss the personalized recommendation based on the customer **knowledge acquisition**, which is a direction of application of this research. In the future research, to improve the efficiency of customer demand **knowledge acquisition**, we should integrate the method proposed in this chapter with telecom operator's OSS (Operation Support Systems) using business intelligence tools. Then, the research result in this chapter can play a more significant role in the telecommunication operation management.

ACKNOWLEDGMENT

The research work in this chapter was supported by the National Natural Science Foundation of China (Project No.: 70701005), the Specialized Research Fund for the Doctoral Program of Higher Education of the People's Republic of China (Project No.: 20070013014), and the Open Research Fund between Beijing University of Posts and Telecommunications and IBM.

REFERENCES

Boulton, R., Libert, B., & Samek, S.M.. (2000). Cracking the Value Code: How Successful Businesses Are Creating Wealth in the New Economy. *Harper Collins Publishers Inc*, New York.

Burs, M. J., & Woodruff, R. B. (1991). Value: An Integrative Perspective. *Proceedings of the Society for Consumer Psychology*, (pp. 59-64).

Chai, H. Z. (2006). Personalized Information Recommendation Service. *Journal of information, 9*, 118.

Chien, H., & Wu. S. (2002). SODPM: a Sequence-oriented decision processmodel forunstructured group decision problems. *Behaviour& Information Technology*, *21*(1), 59-69.

Inoue, A., Takahashi, S., & Nishimatsu, K. (2003). Service Demand Analysis Using Multi-Attribute Learning Mechanisms. *KIMAS*, IEEE, (pp. 634-639).

Keryer, T.N. (2001). I-mode: a successful launch of the mobile Internet market. *Alcatel Telecommunications Review*, (pp. 64-67).

Li, F., & Whalley, J. (2004). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy*, *26*(9), 451-472.

Liou, Y. I. (1990). Knowledge acquisition: issues, techniques, and methodology. Paper presented at *Proceedings of the 1990 ACM SIGBDP conference on trends and directions in expert systems*, Orlando, Florida, United States: ACM Press, (pp. 212–36).

Mark, H., Catja, P., & Peter, A. (2004). Using Means-End Chains to Build Mobile Brand Communities. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences. Track7, 7*, 70196.3. Washington: IEEE Computer Society.

Meier, H., Friedrich, D. R., & Blankensteir, H. (2003). Mobile data: creating sustainable value. *Booz|Allen|Hamilton*, (pp. 1-12).

Pikkarainen, H. (2001). 3G market profile: existing services for 3G. *Nokia Networks*, (pp. 1-19).

Qi, J. Y., Wu, F., & Shu, H. Y. (2007). AI techniques in telecommunications industry. *Expert Systems*, September 2007. *24*(4), 271-291.

Reynolds, T., & Gutman, J. (1988). Laddering Theory, Method, Analysis, and Interpretation, *Journal of Advertising Research*. *28*, 11-29.

Sharp, B., & Sharp, A. (1997). Loyalty Programs and Their Impact on Repeat-purchase Loyalty Patterns. *International Journal of Research in Marketing, 14,* 473-486.

Simpson, P. M., Siguaw, J. A., & Baker, T. L. (2001). Value Creation Supplier Behaviors and Their Impact on Reseller-Perceived Value. *Industrial Marketing Management*, *30*, 119-134.

Srinivasan, S.S., Andeson, R., & Ponnavolu, K. (2002). Customer Loyalty in E-commerce: an Exploration of Its Antecedents and Consequences. *Journal of Retailing*, *78*, 41-50.

Stienstra, M., Baaij, M., & Bosch, F. V. D. (2004). Strategic renewal of Europe's largest telecom operator (1992-2001): from herd behavior towards strategic choice? *European Management Journal*, *22*(3), 273-280.

Walker, B. A., & Olson, .J. C. (1991). Means-end chains: Connecting products with sefl, *Journal of Business Research*, *22*(2), 111-118.

Wan, Y. H., Qian, Y. Y., & Li, J. (2005). The Analysis of Customer Value Dimensions' Difference of the Bank's Interacting Channel, *The Second IEEE Conference on Service Systems and Service Management*, (pp. 193-198).

Wang, H. L., & Feng Y. Q. (2007). Decision support system facing unstructured decision making based on decision flow. *Journal of Harbin Institute of Technology, 10*, 1638-1641.

Wang, S., & Archer, N. (2007). Electronic marketplace definition and classification: literature review and clarification. *Enterprise Information Systems*, *1*(1), 89-112.

Zhang, A., Melcher**,** A., & Li, L. (2004). Mapping the relationships among product complexity, information technology, and transaction governance structure: A case study. *Journal of Management Systems*, *16*(4), 41-55.

KEY TERMS

Artificial Intelligence (AI): Artificial Intelligence is defined as the science of artificial simulation of human thought processes with a computer.

Customer Value: Customer value is a perception of what a customer wants to accomplish with a help of products, in order to reach a desired goal. Customer value describes the nature of the relationship between the customer and the products.

Customer Value Hierarchy: In this chapter, we define that customer value hierarchy consists of the customer demand objective layer, the consequence layer and the attribute layer.

Data Mining: It is the process of extracting interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from a huge amount of data.

Knowledge acquisition (KA): It is a process of identifying, eliciting, and verifying or validating domain-specific knowledge.

Personalized Commendation: According to customer's demand, interest or behavior mode, actively commend some individual products, services or information to customers that who are interested in those commodities.

Structured Decision-Making: The structured decision-making is an ordinarily repeated decision-making with some rules. Enterprises can make an arrangement in advance to realize their anticipative results or object.

Chapter XXIV
Telecommunication Customer Detainment Management

Jiayin Qi
Beijing University of Post and Telecommunications, China

Yuanquan Li
Beijing University of Post and Telecommunications, China

Chen Li
Beijing University of Post and Telecommunications, China

Yingying Zhang
Beijing University of Post and Telecommunications, China

Jing Tan
Beijing University of Post and Telecommunications, China

ABSTRACT

This chapter proposes an integrated methodological system of telecommunication customer detainment management, including telecommunication customer churn prediction and strategy formulation of customer detainment management. The formulation of churn customer detainment management strategy includes customer detainment value assessment, customer detainment level determination, enterprise-attribution approach based on customer detainment strategy analysis, evaluation and implementation, and so forth. Future research in this field is discussed at the end of this chapter.

INTRODUCTION

With the increasing competition in the telecommunication market, customer-winning battles between telecom operators are becoming fiercer and fiercer. In order to gain more market share, telecom operators continually launch and offer new products and services. However, this increases the volatilities of customers to a large extent, leading to the frequent occurrences of **customer churn**.

A report from Gartner's survey reveals the cost of developing a new customer is usually 4-5 times greater than that of retaining an old customer. Another study shows that, with reducing customer churn rate by 5%, a company's profit can increase by 25%-85%. Hence, it is obvious that the loss of a large number of customers can lead to huge losses of the operators.

According to the statistics, in 2003, the average telecommunication customer churn rate in the United States was 30%, while the one in Europe was 25%, and in Asia 48%. The churn of customers would cause huge losses to the telecom operators. Firstly, it accounts for direct losses of the operators' profits; secondly, it makes the costs spent during the development periods of the churn customers become futile; finally, it has a negative impact on the images of the operators. So, it is becoming common knowledge in business, that retaining existing customers is the best core marketing strategy to survive in industry. (Kim, et al., 2004)

In recent years, as the international telecommunication market gradually has become saturated and the increment of new customers has slowed down, maintaining and retaining customers has become the focus of attention more than ever. How to actively develop new customers and at the same time reduce the churn rate of old customers has received extensive attention from both domestic and foreign major telecom operators. At the same time, a large scale implementation of information systems has provided support for computerized customer detainment management for telecom operators. The predictions of behavior, customer value, customer satisfaction and customer loyalty are examples of some of the information that can be extracted from the data that should already be stored within a company's database (Hadden et al, 2005).

Under this background, customer detainment has been paid attention to and widely adopted by the telecom operators.

BACKGROUND

According to the telecommunication customer life cycle theory, telecommunication customer detainment management is the key management task in the declining stage of customer relationship. (Shu & Qi, 2004) Telecommunication customer detainment management includes identifying whether a customer has entered a declining stage; once the customer has entered a declining stage, re-defining customer relationship and re-establishing customer relationship to start a new cycle of customer relationship with the customer if the customer is worth detainment; or stopping investing resources in unrecoverable customer relationship and terminating the relationship in appropriate ways. In other words, identifying potential churn customers and segmenting them, formulating, evaluating and optimizing targeted customer detainment strategies are the main tasks of telecommunication customer detainment management. Through these tasks, the telecom operators can detect churn tendencies of customers and then take effective measures to win potential churn customers back or terminate customer relationships with them in appropriate ways.

Though customer detainment management is very important in telecom operators' daily work, most studies are mainly conducted on constructing **customer churn** prediction system, while rarely being conducted on selecting detainment targets, formulating detainment strategies, detainment effect evaluation and so on, even more rarely on integrated methodological systems of telecommunication customer detainment management. From the 1990s to date, a large number of international journals and the international conferences have focused on designing churn prediction models and algorithms by using data mining technologies to construct more effective churn prediction models (Lian & Richard, 2004; Kristof & Dirk, 2006; Au,& Yao , 2003; Wei & Chiu, 2002; Bloemer et al., 2002; Mozer et al., 2000; Ng & Liu, 2001; Wei & Chiu, 2002), while only a few researches have been done on other aspects of customer detainment management. Shu and Qi proposed a methodological system in regard to telecommunication customer detainment management in "Telecommunication Customer Life Cycle Management" (2004). However, customer detainment level was not investigated in this system. Churn management efforts should not be made across the entire customer base because (i) not all customers are worth retaining, and (ii) customer detainment costs money; attempting to retain customers that have no intention of churning will waste resources. (Hadden et al., 2005) Under this background, the telecom operators have to set different priorities to customers so that they can detain the customers of high importance first. Hence, it is necessary to develop an integrated methodological system of telecommunication customer detainment management to achieve a better customer detainment by the operators.

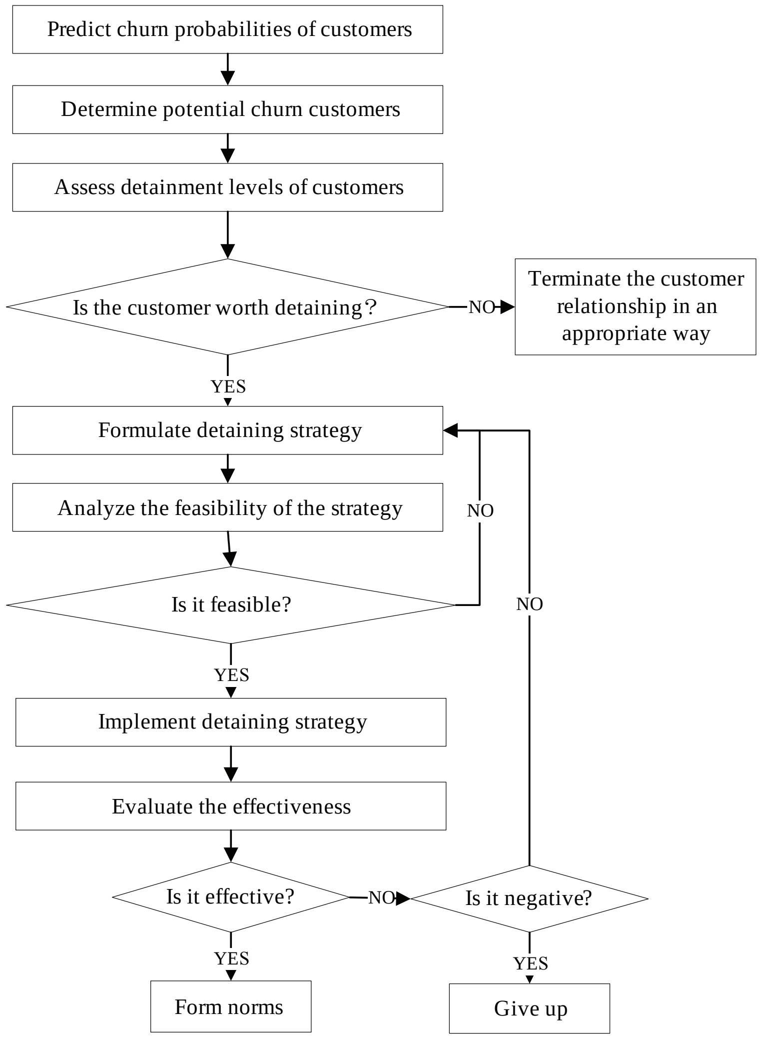
This chapter is organized as follows: Section 3 presents the overall process of telecommunication customer detainment management, Sections 4, 5, 6, and 7 respectively discuss **customer churn** prediction modelling, customer detainment level assessment, detainment strategy formulation, detainment strategy evaluation, and continuous optimization, which are main phases in customer detainment management process; Section 8 describes some commonly used customer detainment methods; Section 9 draws the conclusion of this chapter and provides future research directions.

DESIGN OF TELECOMMUNICATION CUSTOMER DETAINMENT MANAGEMENT PROCESS

An integrated customer detainment solution includes five parts. These five parts execute in a linear or sequential manner in the following order: **customer churn** prediction, customer detainment level assessment, detainment strategy formulation, detainment strategy implementation, detainment effect evaluation and adjustment of models and strategies, as shown in Figure 1:

I. **Customer churn** prediction: Build up customer churn prediction models to analyze the churn probabilities of current customers and detect potential churn customers.
II. Potential churn customer determination: **Churn prediction model** only estimates churn probability of each customer, without telling which customers will be potential churn customers and which ones will not. The task of potential churn customer determination is to determine a churn probability threshold value p_0, with which the potential churn customers and regular customers can be determined.
III. Customer detainment level assessment: In fact, not all customers with high churn probabilities are worth detaining. In order to utilize the resources in effective ways and enhance the detainment effect, it is necessary to segment customers according to churn probabilities and detainment values, and then set high detaining priorities to the ones with both high churn probabilities and high detainment values. For those who are unrecoverable or with low detainment values, appropriate measures to terminate customer relationships are needed.

Figure 1. Telecommunication customer detainment management process



IV. Detainment strategy formulation: After the last two steps, customers who are worth detaining have been selected from potential churn customers. However, the number of the selected customers is always large and it is difficult to analyze each customer and formulate corresponding detainment strategy for each one. In this case, it is helpful to subdivide the selected customers into different types and then formulate corresponding detainment strategy for each type.

V. Detainment strategy feasibility analysis: To evaluate if the detainment strategy is feasible and can benefit the company or not, an analysis of detainment strategy feasibility is needed. To conduct this, both the long/short term revenue and cost should be taken into account. If the strategy is identified to be feasible, it can be implemented afterward; if not, further improvement on the strategy is required.

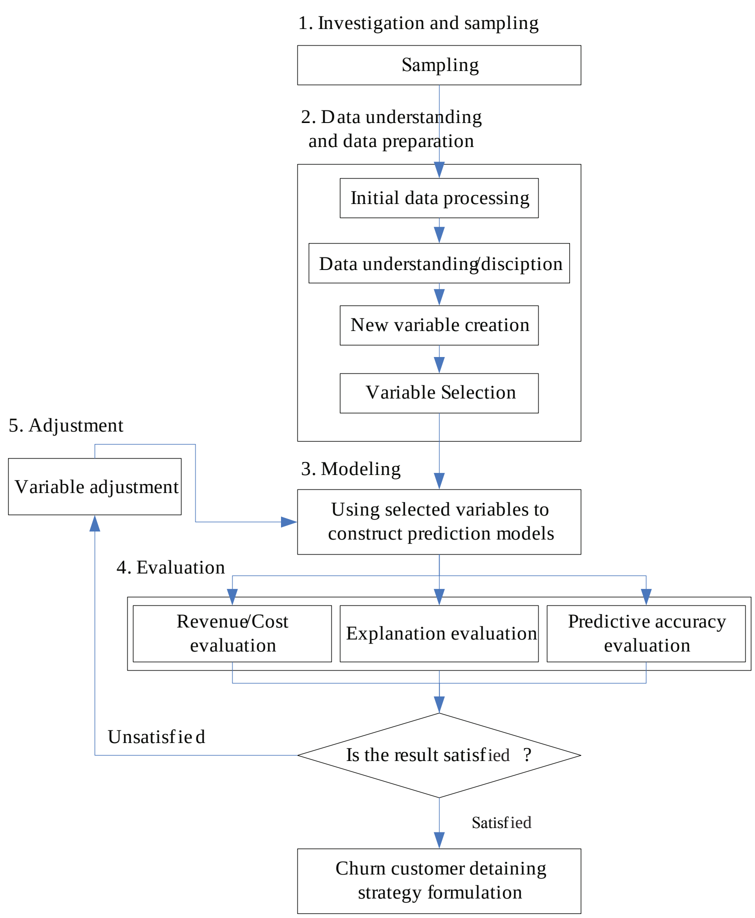
V.I Detainment effect evaluation and strategy optimization: If the percentage of retained customers and the average consumptions of customers keep increasing in one/two/three months after implementation of a detainment strategy, the strategy is proved to be effective; otherwise ineffective.

For those who are unrecoverable and with low detainment values, appropriate measures should be taken to terminate the customer relationships. The measures are not discussed as main contents in this chapter, since they are mostly about practical skills.

TELECOMMUNICATION CUSTOMER CHURN PREDICTION

The core of telecommunication **customer churn** prediction is to construct effective churn prediction models to detect customers with high churn probabilities or in the **customer relationship decline phase**, which is the first step of customer detainment. Telecommunication customer churn prediction models apply data mining technologies to build the mathematical models which use telecommunication customer characteristics (e.g., customer basic information, bill information, call details, service logs, contractual information and credit information, etc.) to predict **customer churn** probabilities. Since telecommunication customer churn prediction employs data mining technologies to construct prediction models, its process can be categorized into five phases by referring to the data mining process, as shown in Figure 2:

Figure 2. Telecommunication customer churn prediction process



Investigation and Sampling

Different telecom operators may have different definitions of churn customer. Customer data storage by different operators may also vary. Hence, model builders should investigate the telecom operators comprehensively before modeling, in order to get clear definitions of churn customers and positive customers. Builders should understand the customer condition of available data in detail, so that a feasible and practical data sampling solution can be proposed in the next step. Usually, a data sampling solution should depict the size of the sample, the list of fields (attributes) to be extracted, the time scale of data, and the sampling proportions of actual churn customers and regular customers.

Data Understanding and Data Preparation

Data understanding and data preparation mainly include four phases: initial data processing, data understanding/description, new variable creation and variable selection. The task of initial data processing is to determine and collect initial data; the task of data understanding/description is to describe the data which has been acquired, and then conduct initial data exploration and examine the quality of the data; new variable creation is to process the initial data, including selecting data for prediction model, cleaning and transforming data, as well as integrating and reducing data for modeling tools; variable selection is to select variables which are useful for the prediction as inputs of churn prediction models, from the variables generated from the last step.

Since the integrated discussion on variable selection is rarely found in previous studies, methods for variable selection are hereby discussed.

I. Significance of Variable Selection

The principle of variable creation is to make the most useful information for **customer churn** prediction out of the customer data. Following this principle, the number of the variables is bound to be large. There could be positive variables and negative variables for prediction. Using all variables as inputs to train churn prediction models can bring burdens to the training process of models. Hence, the variable selection is a quite important step.

II. Principles of Variable Selection

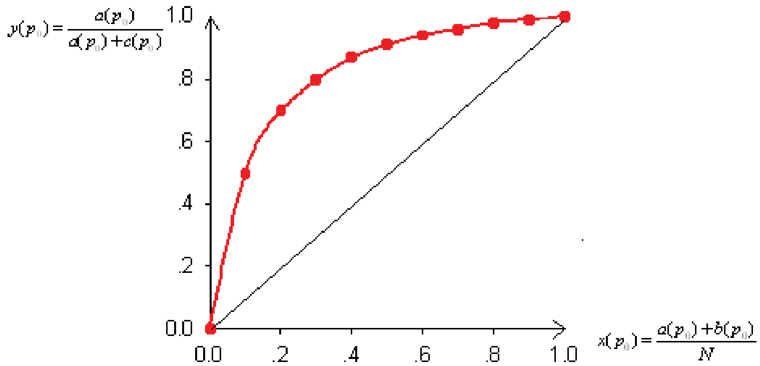
As aforementioned, the variable creation means to extract potential **customer churn** information as much as possible. So this information (large amount of variables) unavoidably contains lots of overload information. Input-feature selection achieves both data cleaning and data reduction by selecting important features and omitting redundant, noisy, or less informative ones (Yan et al., 2003). So, it is necessary to extract useful and brief ones from the variables.

Principle 1: The Classifying Abilities of the Variables Should be High

The classifying ability here means the ability of a variable to classify/predict churn customers from regular customers. Many researchers have paid much attention to this problem: It has been suggested that neural networks can be used for feature selection (Meyer & Watzel, 1998). Ng and Liu have performed feature selection by running an induction algorithm on the dataset.(Ng & Liu, 2001) A method suggested by Datta et al. involves initially finding a subset of features from the data warehouse by manually selecting those that appear most suitable for the task.(Datta et al., 2001)

We use the AUC (Area Under ROC Curve) method to measure the predictive ability which was proposed by Yan et al. (Yan & Wolniewicz, 2004).

Figure 3. An example of ROC curve



AUC (Area Under ROC Curve) Method

AUC is the area between an ROC (Receiver Operating Characteristic) curve and the X- axis. According to the definition of **ROC curve**, $y(p_0)$ is the sensitivity of the model for a given probability threshold value p_0. The sensitivity is a measure of accuracy for predicting target *A* which is the ratio of the number of customers correctly predicted as *A* to the number of total actual customers in *A* under threshold value p_0. $x(p_0)$ is the ratio of the number of customers predicted as A to the number of total customers for a given churn probability threshold value p_0. An example of **ROC curve** is shown in Figure 3.

Where *a* denotes the number of the customers who are churners and are predicted as such; *b* denotes the number of the customers who are not churners but are predicted as churners, *c* denotes the number of the customers who are churners but are predicted as not churners and *d* denotes the number of the customers who are not churners and are predicted as such. Since *a*, *b*, *c* and *d* are influenced by the threshold value p_0, we express them as $a(p_0,)$ $b(p_0,)$ $c(p_0,)$ $d(p_0)$. *N* is the size of the test set.

ROC curve is a graphical display that gives the initiative measure of the classifying accuracy. If the **ROC curve** in the lower left corner has a steep upward trend, it means the prediction model has a high sensitivity even with strict selection criteria. Thus the model is proved to have high accuracy. The closer the **ROC curve** is to the upper-left, the higher classifying accuracy the model is. The area under **ROC curve**-AUC is a frequently used index to evaluate the classifying accuracies of models.

AUC can be calculated in the form

$$U = \frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} I(x_i, y_j)}{mn} \tag{1}$$

Where,

$$I(x_i, y_j) = \begin{cases} 1 \; if\,(x_i > y_j) \\ 0 \text{ otherwise} \end{cases}, \tag{2}$$

$\{x_0, x_1,..., x_{m-1}\}$ is the set of different predicted values for churn customers, while $\{y_0, y_1,..., y_{n-1}\}$ is the one of different predicted values for positive customers.

Principle 2: Mutual Information Between Variables should be Relatively Low

Mutual information is a concept to measure how much information one variable tells about another one. A higher **mutual information** between two variables means the two variables share more similar information. In this case, we can delete some redundant variables, to make sure the **mutual information** between variables is relatively low.

Mutual information is computed by:

$$I(T, a) = H(T) - H(T / a) \quad (3)$$

Where:

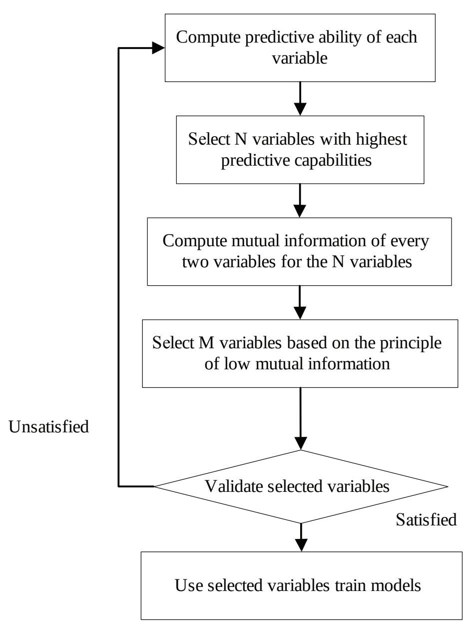
$$H(T) = -\sum_{i=1}^{n} p(a_i) \log p(a_i) \quad (4)$$

Is the entropy of T

$$H(T / a) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p(a_i b_j) \log p(a_i / b_j) \quad (5)$$

Is the conditional entropy of T, given the value of a;

Figure 4. Variable selection process



The former one reflects the amount of "disorder" of the target variable *T*, while the latter one reflects the amount of "disorder" of target variable T after considering *a*. Hence, **mutual information** is the amount of information that *a* provides to classify *T*.

III Variable Selection Process

Variable selection process is shown in Figure 4:

In practical applications, a variable is considered to be useful for churn prediction only when its AUC is larger than 0.5. In this way, the number of the variables can be determined. Two variables are kept only when the **mutual information** between the two variables is smaller than 0.5 and then the number of the variables used for building up prediction models can be determined. The variables selected can be used as inputs for initial churn prediction models. If the performances of the models are satisfying, the variables could be the final variables for prediction, otherwise future adjustment of the cutoff values of AUC and **mutual information** should be made to get better performances.

Modeling

Customer churn prediction model is based on two basic assumptions:

Assumption I. Different customer groups have different churn tendencies;

Assumption II. Certain customers who churned from a company will have certain unusual consuming behaviors.

Customer churn prediction model is eventually fitted into a churn probability function $P(X)$, in which X stands for the eigenvector of customer (includes information stored in database which reflects the churn tendency, e.g. demographic information, account information and customer calling behavior). $P(X)$ is defined within [0, 1] and describes the **customer churn** rate if there is no detain measures for this customer in a certain coming period. Many methods are used for fitting $P(X)$. The most common method is the multiple regressions, neural network, decision tree and the combination of these 3 methods. What needs to be mentioned is there is so few churned customers for a company that the ratio of regular customers to churned customers is high skew distribution in the whole sample. This point must be taken into fully consideration when modeling and feasible measures must be used.

Model Evaluation

Prediction and explanation are two main functions of churn prediction model. Issues of resource allocation, always somewhat thorny, still bring more challenges. Through prediction, the enterprise can focus on a small scale of customers (i.e. the most likely churners), and then pool resources to conduct **customer detainment** with pertinences. Through explanation, the enterprise can find some regular patterns of customers' behavior, which will help the enterprise to better understand which factors have great impacts on customers' loyalty and churn tendency. As for prediction, response rate, captured rate, lift value, ROC value and revenue curve are the common indexes to evaluate the prediction effect of the churn model.

Response rate indicates that the percentage of churned customer in this situation:

I. Use **churn prediction model** to get the churn possibility of all customers;
II. Sort all customers by churn possibility;
III. Within a certain percentage of all customers (e.g. 10%), response rate stands for the percentage for which real churned customers account.

Captured response means the percentage that after sorting all customers by predicted churn possibility, how many real churned customers within a certain range of customers (e.g. 10%) account for in all real churned customers.

Lift value indicates the ratio of real churned customers within top n (n=10, 20…100) customers who has been sorted by churn possibility according to the prediction model to real churned customers within top n (n=10, 20…100) customers who has been sorted by churn possibility according to subjective judgment.

That all aforementioned indicators are higher stands for a better model prediction performance.

Model Adjustment

If the model evaluation provides a satisfied result, there is no need adjusting the model and **customer detainment** strategy should be the next step; if not, the model should be adjusted or some other models should be selected.

IDENTIFICATION OF CHURN CUSTOMERS

Normally, the assertion for classification of customers is not provided by prediction models. Models return a probability from 0 to 1 to say customer belonged to a certain class rather than a Yes or No conclusion to answer whether customers will churn. Model users should take all constraints (e.g. different resources of the company) into consideration, set a threshold for churn probabilities accordingly and finally classify customers. A high probability stands for a high possibility it belongs to a class. So there should be a probability threshold p_0, if $p \geq p_0$, this customer is identified as class A and vice versa.

Before implementing the detainment strategy, customers should be identified for detainment according to the company's practical situation, e.g. to classify into potential churned customer class. In this chapter, churn probability threshold is set by the maximum earning principle.

Table 1. Model prediction matrix

	Real churned	Regular
Predicted churned	$N_{11}(p_0)$	$N_{10}(p_0)$
Predicted regular	$N_{01}(p_0)$	$N_{00}(p_0)$

Model Prediction Matrix

No model could achieve perfect predictions and wrong predictions are always in place. So prediction results could be classified into 4 types: real churned is predicted churned, real churned is predicted regular, regular is predicted regular, regular is predicted churned. The numbers of each type are N_{11}, N_{10}, N_{01} and N_{00}, as illustrated in Table 1.

Notes: p_0 is the threshold value which is used to judge churned or not. If churn probability is bigger than p_0, this customer could be identified as potentially churned. Accordingly, for different p values, N_{11}, N_{10}, N_{01} and N_{00} should have different value.

Cost/Profit Matrix

When conducting churn prediction, simply higher prediction accuracy could not make sure the highest profit for the company. Because profits are different between predicting churn customer accurately and predicting regular customer accurately in customers churn prediction.

Table 2. Cost/Profit matrix

	real churned	regular
Predicted churned	Cost / Profit	Cost / 0
Predicted regular	Cost / - Profit	- Cost / 0

Based on the aforementioned hypothesis, we could get Cost/Profit Matrix as seen in Table 2:

Identification of Churned Customer Group

Based on Model Prediction Matrix and Cost/Profit Matrix, total profit for detainment strategy is as seen in equation (6):

$$
\begin{aligned}
&Total\ Profit(p_0) \\
&= (Profit - Cost) \times N_{11}(p_0) + (0 - Cost) \times N_{10}(p_0) + \\
&\quad (-Profit + Cost) \times N_{01}(p_0) + (0 + Cost) \times N_{00}(P_0)
\end{aligned} \tag{6}
$$

Therefore, for each threshold value p_0, there is different total profit. When maximizing this equation, a threshold value p_0 is obtained. Based on this p_0, a churned customer group is identified for detainment strategies.

DETAINMENT RESPONSE HIERARCHY ANALYSIS ON CHURNED CUSTOMER

The formulation of **customer detainment** strategy is largely based on detainment response hierarchy of churned customers. And this hierarchy is related to a variety of factors. Among these factors, customer churned tendency and **customer detainment value** play a major role. Customer churned tendency is measured by customer churned probability from the churn prediction model. A high churned probability indicates a high churned tendency. The major consideration of companies should be the combination of churned tendency and detainment value to achieve a best detainment result with the help of efficient usage of limited resources when designing detainment strategies.

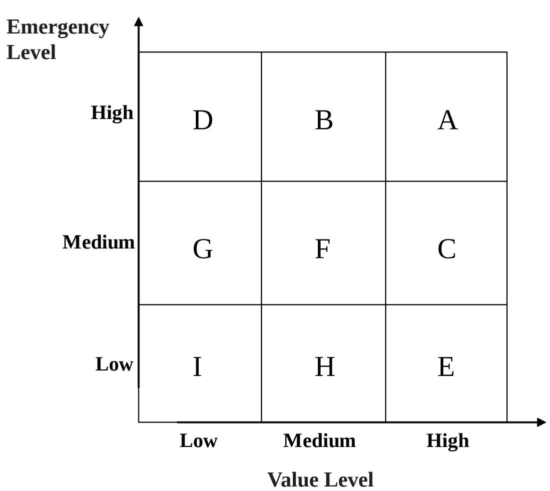
Churned Customer Value Analysis

This part will focus on churned customers value as customer churned tendency has been discussed before. Detainment value evaluation is based on the evaluation of a single customer's consuming behavior. It is an indirect evaluation method without evaluating directly potential profit from the customer or evaluating directly the cost of detaining the customer. In this method, customer's consuming behaviors before he churns have been reviewed comprehensively to evaluate this customer's value. It is based this hypothesis: if a customer has a strong profitable and stable consuming behavior before he churns, this customer could be a high-quality customer and is worth to detain. Indicators that follow are used to analyze the **customer detainment value**:

- Customer average telecommunication consumption amount
- Customer telecommunication consuming stability

Based on these two value indicators, **customer detainment value** could be evaluated comprehensively.

Figure 5. Customer segments and customer detainment levels



Detainment Response Hierarchy Analysis on Churned Customer

Companies are restricted by cost as well as other factors when they try to detain customers. Normally, companies need to evaluate the detainment response level of customers to decide the priority of implementing **customer detainment** strategies. In this way, company could optimize the allocation of resources. In this research, the main thought is to consider both **customer detainment value** and customer churned tendency, which will provide foundation for designing **customer detainment** strategy.

According to customer churn probability (emergency level) and **customer detainment value** (value level), we divide all potential churn customers into nine customer segments (named from A to I), showed as Figure 5.

Furthermore, in order to determine the priority of each customer segment, three additional criterions are proposed, which are:

Criterion 1: The higher the **customer churn** probability is, the higher the detainment priority is, on condition of the same **customer detainment value**;

Criterion 2: The higher the **customer detainment value** is, the higher the detainment priority is, on condition of the same **customer churn** probability;

Criterion 3: The priority of the **customer churn** probability is higher than **customer detainment value**. For example, the priority of D (high **customer churn** probability and low customer detainment value) is higher than E (high customer detainment value and low **customer churn** probability).

Based on these criterions, priorities of nine customer segments detainment level can be determined at last, that is: A>B>C>D>E>F>G>H>I.

DESIGN CUSTOMER DETAINMENT STRATEGY

Customer churn reason analysis is the first step for designing telecommunication **customer detainment** strategies. When the decline of customer relationship happens, the **enterprise attributive analysis** will be used to analyze the declining reason for designing customer detainment strategy.

Enterprise attributive analysis has these steps:

Step 1: To identify **customer churn** reason. Separately analyzes the declining reasons of customer relationship on different detain hierarchy.

Step 2: To do **enterprise attributive analysis.** In an enterprise's point of view, the controllability and the frequency of declining reason are bases for judging that whether the declining reason is necessary to improve, whether it can be improved, and how much it should be improved. The analysis processes are shown in Figure 6 and Figure 7.

Figure 6. Enterprise attributive analysis

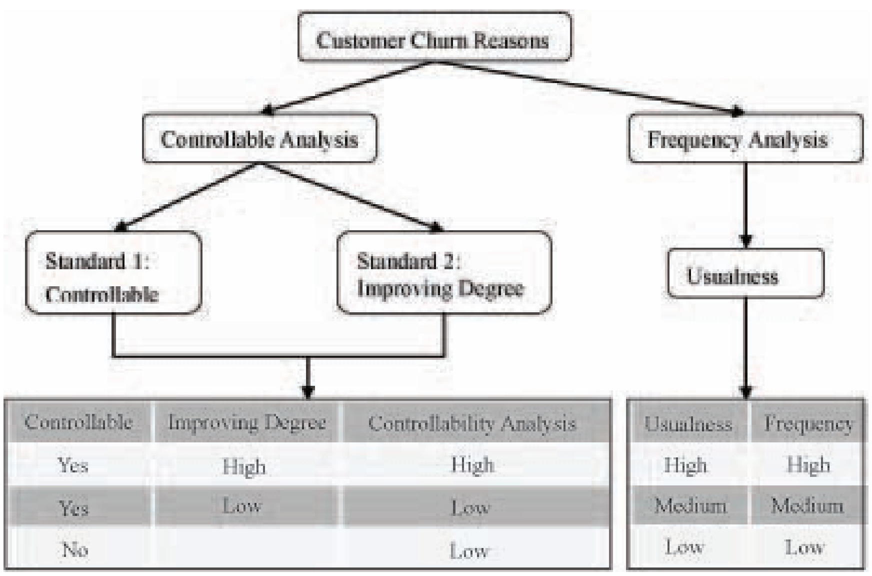


Figure 7. The rules of determining controllability and frequency

Controllable	Improving Degree	Controllability Analysis
Yes	High	High
Yes	Low	Low
No		Low

Usualness	Frequency
High	High
Medium	Medium
Low	Low

Figure 8. Classification matrix

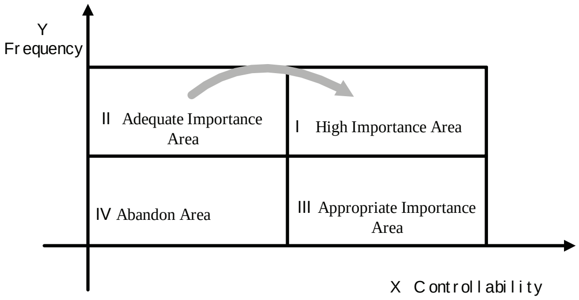
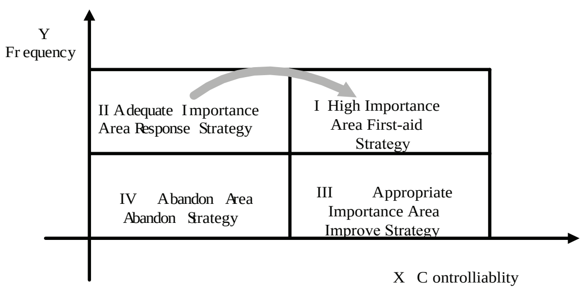


Figure 9. Strategies in quadrants



Step 3: Controllability and frequency analysis. Matrix of controllability and frequency is build to identify churn reasons. The value and position of churn reasons are analyzed quantitatively in this matrix. See Figure 8.

Step 4: To formulate different detainment strategy on different churn customer segments. In this step, different strategies will be designed for 4 categories of customers that in the 4 quadrants separately. See Figure 9.

First-aid strategy: intensify the controlling power, lower the frequency and recover customer relationship as soon as possible.

Response strategy: try to transform low-controllable to high-controllable, regain customers and win the initiative in competition.

Improve strategy: improve all related factors, lower the frequency to a low level, try to regain customers and make them useful for the development of companies.

Abandon strategy: choose a right time to terminate customer relationship, release the resource and try to provide a good impression for customers.

FEASIBILITY ANALYSIS FOR CUSTOMER DETAINMENT STRATEGY

Questions about whether **customer detainment** strategy is feasible and whether detainment strategy will bring profit are answered by feasibility analysis. The long-term and short-term profit and cost are taken into comprehensive consideration. If the detainment strategy is feasible, the enterprise can implement this strategy; if not, this strategy should be improved or abandoned.

On the side of profit, it firstly comes from the direct revenue after successful detainment. In other words, the income comes from those detained customers. The second source of revenue is the indirect revenue, including profit from new customers who are attracted in the detainment process, from avoiding the potential loss of churned customers, and from the invisible revenue from the implementation of strategies, e.g., invisible revenues from better enterprise image and better brand influence.

Detainment cost includes not only the cost of implementing detainment strategy, but also the maintenance cost after successful detainment, such as network occupancy cost and related system building cost.

In addition, the analysis includes some invisible revenue which could not be quantified. When comparing the profit with the cost of the detainment strategies, the setting that enterprise faces should be fully considered. For a developing telecom operator, direct revenue is focused and indirect revenue accounts for a low percentage when implementing detainment strategies. For an operator which just experienced bad social

Chart 1.

- **N_i is the number of customer who belong to type i**
- **P_i(Retrieve) is retrieve rate which indicates the ratio of the successful detained customers number to type i customers number after company has implemented detainment strategy on type i customers.**
- **$ARPU_i$ is the expected $ARPU$ (Average Revenue Per User) value after the type i customer's winning back.**
- **T_i is the period that company could get revenue from the customer after he/she has been successfully detained. Its unit is month.**

IR(Indirect Revenue)

image matters, it may expect a significant amount of indirect revenue by implementing detainment strategy.

(1) Calculation for Direct Revenue

Direct Revenue

Direct revenue is an income that one type of customers (e.g. Type *i* customer) brings to the company after they has been successfully detained.

$$DR_i(Customer) = P_i(Retrieve) \times N_i \times ARPU_i \times T_i \quad \text{[see Chart 1.]} \qquad (7)$$

R(Indirect Revenue) is an income that detainment strategies bring to the company indirectly.

Customer detainment strategies will not only bring profits to operators by improving satisfaction level of churn customers and attracting people surrounding these customers as new customers, but also reduce negative information broadcasted to the public from churned customers, and then reduce the **customer churn** rate and also the potential loss. In addition, indirect revenue also includes the invisible revenue from the improved brand and the corporate image.

Analysis will be conducted in 3 aspects:

① IR_i(*Attractive*) refers to the revenue from the new customers who are attracted by the successful detainment strategy implementation.

Chart 1.

(1) **$P_i(Attractive)$ is the ratio of new customers who are attracted by successful retention strategy implementation to all new customers. Its value is between 0 and 1.**

(2) **$M(Attractive)$ is the average number of new customers per month after implementation of retention strategy.**

(3) **$ARPU_A$ refers to the $ARPU$ value of new customer.**

(4) **T_i is the predicted retention period. Its unit is month.**

Chart 3.

- $P_i(Decrease)$ **is the ratio of reduced number of churned customer due to retention strategy to the all reduced number of all churned customers. Its value is between 0 and 1.**
- $M(Decrease)$ **is the average reduced number of churned customers after retention strategy compared to number of churned customers before retention strategy.**
- $ARPU_B$ **is the** $ARPU$ **value of reduced churned customers.**

$$IR_i(Attractive) = P_i(Attractive) \times M(Increase) \times ARPU_A \times T_i \quad (8)$$

[see Chart 2.]

② $IR_i(Decrease)$ is the revenue from the reduced number of churned customers who are detained by detainment strategy.

$$IR_i(Decrease) = P_i(Decrease) \times M(Decrease) \times ARPU_B \times T_i \quad (9)$$

[see Chart 3.]

Explanation:

$M(Decrease)$ in equation (9) indicates the monthly average reduced number of churned customers after detainment strategy. However, if the monthly average reduced number actually grows because of the marketing efforts by competing rivals or emergencies, $M(Decrease)$ will be negative and the aforementioned equations will not work. While, in such a case, the implemented **customer detainment** strategy still has a positive effect because there will be more churned customers without detainment strategy. So, if it happens, the reduced number of churned customers is required to be confirmed by market research. The revenue from these customers is equal to $M(Decrease)$ multiplied by $ARPU_B$ and $T_{i.}$

③ *Invisible* Revenue (*IR*) is the revenue from detainment strategy which is hard to quantify. This revenue will facilitate company's development in long term. It could be divided into several parts as follows:

$$IR_i(C) = R_i(Brand) + IR_i(Company) \quad (10)$$

[see Chart 4.]

Chart 4.

- C_u means the monthly cost on each customer which is used to maintain the customer relationship, e.g. customer service cost
- N_i indicates the number of churned customers in type i
- $P_i(Retrieve)$ is the retrieve ratio of the number of detained customers in type i to the total customer number of type i after implementing retention strategy on type i customers.
- T_i is the predicted retention period. Its unit is month.

Chart 5.

- $C_i(Strategy)$ **is the strategy cost which is mainly from strategy itself, e.g. cost of mobile phone when mobile phone is free by promotion strategy or discounted cost in discounted strategy**
- $C_i(Advertisement)$ **is the advertisement cost which is used to combine with retention strategy to improve the advertising effects**
- $C_i(Human)$ **is the human resource cost which is used to pay for the extra cost to recruit staff to do market investigation, analysis, advertising and communication with customer**
- $C_i(Agency)$ **is the agency cost which is used to pay for the commission for agencies. These agencies are hired to deal with certain part of retention strategy.**
- $C_i(Channel)$ **is channel cost which is used to manage the distribution channel of retention strategy.**

Once a customer churned, company should actively take some strategies to win him/her back, e.g. more services and preferential treatment. Through communications, customers will know more about the brand and get rid of previous misunderstanding. When these customers spread their new understandings of brand and good image to public, recognition of the brand and good image will come, and it will finally help popularize the brand and attract new customers.

(2) Calculation of Detainment Cost

① Implementation Cost

$C_i(Implement)$ refers to the implementation cost which is the major cost in implementation of detainment strategies. It will be discussed in several aspects as follows:

② Customer Maintaining Cost

Customer maintaining cost is used to maintain the customer relationship with successful detained customer and enterprise within predicted retention period after the customer has been regained by enterprise.

$$C_i(Maintain) = N_i \times P_i\,(Retrieve) \times C_u \times T_i \quad (12)$$

[see Chart 5.]

③ Network Resource Cost

$C_i(Network)$ is a network resource cost which equals the average cost per customer for the whole mobile network multiplied by the number of successful detained customers.

④ Incline Cost

$C_i(Incline)$ is in allusion to the customers who have not churned but tend to churn because they will get more profits through the enterprise's detainment strategies than maintain the relationship with the enterprise all the time. It will lead to certain quantities of churned customers, and brings some loss to enterprise. This loss is called incline cost. e.g. discounted price and promotions. Incline cost is the loss from these temporary churned customers.

⑤ IT Cost

IT Cost is used to build the IT system to maintain customer data and regain customer relationships due to the implementation of detainment strategies. It is noted as $C_i(IT)$.

In summary, the equation will be:

Customer Detainment Value
= Detainment Revenue –Detainment Cost
= Direct Revenue + Indirect Revenue – Implementation Cost – Customer maintaining cost – Network Resource Cost – Incline Cost – IT Cost
= Direct Revenue + Revenue from the new customers + Revenue from the reduced number of churned customers + Invisible Revenue – Implementation Cost – Customer maintaining cost – Network Resource Cost – Incline Cost – IT Cost

(13)

Or:

$$CustomerDetainingValue = DR_i(Customer) + IR_i(Attractive) + IR_i(\text{Decrease}) + IR_i(Brand) + IR_i(Company) - C_i(\text{Im}plement) - C_i(Maintain) - C_i(Network) - C_i(Incline) - C_i(IT) \quad (14)$$

(1) Model-Based Feasibility Analysis for Customer Detainment Strategy

Based on **customer detainment value** equation (14), customer detainment value can be calculated and used for reasonability and feasibility analysis.

Case 1: Direct Revenue > Detainment Cost

$$IR_i(Company) > C_i(\text{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline) + C_i(IT) \quad (15)$$

In this case, detained customers will bring enough direct revenue to counteract all the detainment cost. Furthermore, there is some indirect revenue. So this strategy will of course make more benefit. It has been proved to be feasible.

Case 2: Direct Revenue < Detainment **Cost, but Direct Revenue + Revenue from the new customers + Revenue from the reduced number of churned customers > Detainment Cost, which means**

$$IR_i(Company) > C_i(\text{Im}plement) + C_i(Network) + C_i(Incline) + C_i(IT) \quad (16)$$

But,

$$DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) > C_i(\text{Im}plement) + C_i(Maintain) + C_i(\text{Network} + C_i(Incline) - C_i(IT) \quad (17)$$

In this case, direct revenue could not counteract all the detainment costs, but the revenue that direct revenue plus revenue from the new customers and revenue from the reduced number of churned customers are bigger than detainment cost. Regarding this result plus the invisible revenue, this strategy is still feasible.

Case 3: Direct Revenue + Revenue from the new customers + Revenue from the reduced number of churned customers < Detainment Cost, but the result of Direct Revenue + Revenue from the new customers + Revenue from the reduced number of churned customers + Invisible Revenue - Detainment Cost is unknown, which means

$$DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) < C_i(Implement) + C_i(Maintain) + C_i(Network) + C_i(Incline) - C_i(IT) \quad (18)$$

But

$$DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) + IR_i(\text{Decrease}) + IR_i(Brand) + IRi(Company) - C_i(\text{Im}plement) - C_i(Maintain) + C_i(Network) - C_i(Incline) - C_i(IT) \quad (19)$$

is unknown.

In this case, the visible part in direct revenue and indirect revenue is lower than detainment cost while invisible cost has not been considered. It is required to analyze the feasibility regarding the practical situation.

If there is a strong word of mouth effect, company will suffer a long-term loss from the difficulty in developing brand customers and building image without **customer detainment** strategies. But with strategy, word of mouth effect will facilitate the positive information communication, which helps popularize brand and company's future development. In this case, even though company could not make direct visible revenue in short term, long-term revenue is very high and the strategy is feasible.

Another case is that the word of mouth has a limited impact on people. Though the customer relationship is not recovered, negative information broadcast doesn't have a strong effect. What's more, even if it is recovered, it could not bring high invisible revenue for company. So the strategy is not feasible.

Case 4: Direct Revenue > Implementation Cost, but Direct Revenue + Revenue from the new customers + Revenue from the reduced number of churned customers < Detainment Cost, which means:

$$DR_i(Customer) > C_i(\mathrm{Im}plement) \tag{20}$$

But

$$DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) < C_i(Implement) + C_i(Maintain) + C_i(Network) + C_i(Incline) + C_i(IT) \tag{21}$$

In this case, direct revenue seems to cover the implementation cost and even higher than the implementation cost. Many companies will judge the strategy is feasible. But in fact, these companies neglect the customer occupied network resource and in the long-term, customer maintaining cost and incline cost will grow higher and will make the combination of direct revenue and visible revenue in indirect revenue lower than detain-

Table 3. Strategy feasibility analysis based on customer detainment value

Case	Customer Detainment Value	Feasibility Analysis
1	$DR_i(Customer) > C_i(\mathrm{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline) + C_i(IT)$	Feasible
2	$DR_i(Customer) < C_i(\mathrm{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline)$ But, $DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) > C_i(\mathrm{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline) + C_i(IT)$	Feasible
3	$DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) < C_i(\mathrm{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline)$ But, $DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) + IR_i(Brand) + IR_i(Company) - C_i(\mathrm{Im}plement) - C_i(Maintain) - C_i(Network) - C_i(Incline) - C_i(IT)$	Invisible revenue is required to consider.
4	$DR_i(Customer) > C_i(\mathrm{Im}plement)$ But, $DR_i(Customer) + IR_i(Attractive) + IR_i(Decrease) < C_i(\mathrm{Im}plement) + C_i(Maintain) + C_i(Network) + C_i(Incline) + C_i(IT)$	Invisible revenue, long-term and short-term revenue are required to consider.

ment cost. If invisible cost only accounts for a small percentage of all revenues, this strategy is not feasible practically. And as the time goes by, company will possibly suffer a higher loss.

In this case, company should take practical situation, long term and short term revenue into consideration together to make a decision.

In summary, aforementioned analysis is seen in Table 3.

Detainment Strategy Evaluation and Continuous Optimization

Customer detainment value evaluation is the feedback phase after implementing customer detainment strategies. No matter how successful or not after implementation, company should compare, analyze and evaluate the implementation results with expected goal to find out the reasons for any possible deviation from expected goal and propose guidance and suggestions for detainment response and improvement of detainment strategies. The evaluation will make sure resources are used reasonably and higher economic revenue are gained than attracting new customers, which will together form the closed loop of **customer detainment** management.

The evaluation of **customer detainment** strategies could be monitored by these 2 indicators:

- **Customer Churned Rate**

Customer Churned Rate is the ratio of the customers who transfer from one telecom operator to another or from one brand to another brand in the same telecom operator to all customers. If there is a huge decline after implementations without the affection from other sources, the strategy is proved to be effective. If customer churned rate is still high, the strategy is proved to have some problems and is required to improve.

- *ARPU*

Comparing the *ARPU* of certain groups of customers before and after the implementation of customer detain strategies will show whether the strategies enhance the customer value. If there is a significant rise of *ARPU* without any impacts from other sources, the strategy is proved to have a significant effect on maintaining customer relationships and to enhance customer value. If not, the strategy has the limited effect on company revenue and is required to improve.

While evaluating the **customer detainment** strategy, some other indicators could be taken into account, e.g. ROI (Return of Investment). Customer churned rate and *ARPU* are the two simplest and most practical indicators among all indicators. On one hand, they tell whether strategy is effective on preventing customers from churning; on the other hand, they help judge whether strategy brings more revenue. Therefore, the effect of strategy can be evaluated quantitatively by these 2 indicators. However, customer churned rate are not in line with *ARPU* When customer churned rate reduces, *ARPU* may be lower because of some discount strategy. So company needs to evaluate them comprehensively to realize the strategy effect. What is obvious is that a company does not want to see the situation of rising customer churned rate and reduced *ARPU* happen simultaneously, which indicate the failure of the detainment strategies. On the other hand, the company would love the simultaneous reduced customer churned rate and rising *ARPU*, which indicate the success of the detainment strategies.

FUTURE TRENDS

Competition in the telecom industry will be fiercer because of the deregulation in the global telecom industry, more rapid development of new technologies and services, and convergence of the telecom network, computer network, and television network. Telecom operators' customer management will inevitably transform from single-point static customer management to life circle management. New customer development should focus on high value-added potential customers who are hard to churn. Old customer detainment should focus on high value and stable customers. **Customer detainment** management should focus on detaining high value and stable customers who suffer from the improper **customer detainment** activities. The churned characteristics which are gained from customer detainment management could be used in new customer development to avoid developing customers who are easy to churn. Telecom customer life circle management will be the most important part in telecom customer management in the future. The emphasis of telecom customer life circle management should be on the overall good performance in all phases of the customer life cycle not a perfect performance in only one phase.

In the phase of **customer detainment** management, further research and practices are required. Current research and practice focus mainly on customer churn alarm, and there are only a few studies on customer detainment value evaluation. There are some practical experiences by companies, but research on how to formulate emergent detainment strategy and long-term detainment strategy is lacking. There is also a lack of in-depth research on evaluation methods and resource analysis for detainment strategy. These issues should be the focus of both research and practice in the future.

Regarding our research methodology, aside from the quantitative analysis of **customer churn** alarm, the remaining parts of **customer detainment** management are mainly qualitative. How to apply quantitative analyses to telecom customer detainment management is worth studying. For the management practice, some excellent cases on **customer detainment** are worth sharing and collecting.

CONCLUSION

A comprehensive methodology of **customer detainment** management is proposed in this chapter. This methodology was developed based on the author's management consulting experience with several China telecom operators. Each phase in this methodology has been explained in detail, especially the discussion on the missing points in the current literature.

A good management performance will be gained when this methodology is implemented prudently. Customer detainment management is not the ending phase of customer life cycle management. It should be distilled into the whole customer management processes including new customer development, customer value upgrading, and customer detainment.

ACKNOWLEDGMENT

The research work in this chapter was supported by the National Natural Science Foundation of China (Project No.: 70701005), the Specialized Research Fund for the Doctoral Program of Higher Education of the People's Republic of China (Project No.: 20070013014), and the Open Research Fund between Beijing University of Posts and Telecommunications and IBM.

REFERENCES

Au, W., Chan, C. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*, 532–545.

Bloemer, J., Brijis, T., & Vanhoof, K. (2002). Comparing complete and partial classification for identifying customers at risk. *International Journal of Research in Marketing*, *20*, 117–131.

Datta, P., Masand, B., & Mani, D.R. (2001). Automated cellular modeling and prediction on a large scale, *Issues on the Application of Data Mining*, (pp. 485–502).

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), 2902-2917.

Kim, M., Park, M., & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunications services. *Telecommunications Policy*, *28*, 145–159.

Kristof, C., & Dirk, V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, *Expert Systems with Applications*, *34*(1), 313-327.

Lian, Y., Richard, H. W., & Robert, D. (2004). Predicting Customer Behavior in Telecommunications, *Intelligent Systems*, *19*(2), 50-58.

Meyer-Base, A., & Watzel, R. (1998). Transformation radial basis neural network for relevant feature selection, *Pattern Recognition Letters*, *19*, 1301–1306.

Mozer, M. C., Wolniewicz, R., & Grimes, D. B. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*, *11*(3), 690–696.

Ng, K., & Liu, H. (2001). Customer retention via data mining, *Issues on the Application of Data Mining*. *Artificial Intelligence Review*, *14*(6), 569–590.

Shu, H. Y., & Qi, J. Y. (2004). *Telecommunication Customer Life Cycle Management*. Beijing University of Posts and Telecommunications Press.

Wei, C. P., & Chiu, I. T. (2002). Turning telecommunication call detail to churn prediction: a data mining approach. *Expert System with Application*, *23*, 103-112.

Yan, L., Dodier, R., & Mozer, M. (2003). Optimizing Classifier Performance via an Approximation to the Wilcoxon-Mann-Whitney Statistic. In F. Tom & M. Nina (Eds.), *Proceeding of the Twentieth international Conference on Machine Learning*, (pp. 848-855). Washington: *AAAI Press*.

Yan, L., Wolniewicz, R., & Dodier, R. (2004). Predicting customer behaviour in telecommunications. *IEEE Intelligent Systems*, *19*, 50–58.

KEY TERMS

Churn Alarm Model: The mathematical model or data mining model for predicting customer churned tendency.

Customer Churn: Customer's switching from this services or goods provider to another.

Customer Detainment Management: A management methodology containing customer management process, strategy and evaluation which company uses to prevent customer from churning.

Customer Detainment Strategy: The management strategy which company uses to avoid customer from churning.

Customer Relationship Decline Phase: The last phase of customer relationship life circle management (Recognition Phase, Development Phase, Mature Phase and Decline Phase). When customer has the signs to churn, customer relationship enter Decline Phase.

Detainment Response Hierarchy Analysis: To decide the detainment priority of churned customer in the view of company.

Enterprise Attributive Analysis: It is the analyzing method focusing on customer churned reason in the view of company.

Chapter XXV
How to Plan for an Upgrading Investment in a Data Network?

Frédéric Morlot
Orange Labs, France

Salah Eddine Elayoubi
Orange Labs, France

ABSTRACT

We illustrate how a mobile data network operator can plan an upgrading investment to anticipate explosions of the demand, taking into account the expected profit and the customer satisfaction. The former parameter grows with the demand, whereas the latter sinks if the demand is too high, as throughput may collapse. As the equipment price decreases with time, it may be interesting to wait rather than to invest immediately. We then propose and discuss two methods that help in making the decision. The first one is an actualization algorithm, where the upgrade should be performed when the loss of profit, derived analytically, exceeds the expected discount. The second is a real option-like strategy to hedge against the risk that the investment has to be anticipated. The evaluation of the investment date is then performed, in the latter, by a backward dynamic programming approach, using recent improvements based on least-squares estimations, whereas, in the former, a forward actualizing algorithm is investigated through simulations.

INTRODUCTION

As new multimedia applications are introduced in the market, the traffic is significantly growing in mobile networks. In particular, data flows are growing exponentially, as they have been doing since the middle 90s in the Internet. To face these soaring volumes of data to be transferred, mobile operators must periodically upgrade their equipments to offer higher throughputs and avoid blocking problems. However as the demand does not increase steadily and must be considered as partly random, the upgrade dates are difficult to be forecast (Morlot, et al., 2007).

In this chapter, we consider upgrade investments in a **HSDPA** (*High Speed Downlink Packet*

Access) **cellular network**. Note that, when demand increases, the data communication duration becomes longer and longer for each user until the network is saturated. The individual throughput experienced in the network may become very small.

On the other hand, as the demand rises, the operator increases its profit. When the network starts experiencing saturation problems, throughput and profit may fall. The operator must then upgrade its network by adding new frequency carriers, facing the following trade-off:

- The later the investment, the lower individual throughputs and customer satisfaction. Permanent non-satisfaction will result into churn and additional loss of profit.
- The sooner the investment, the more expensive the costs of upgrade elements.

This chapter aims at modeling analytically the trade-off. We first derive analytical values for capacity, individual throughput and satisfaction as a function of the demand, and use them to calculate operator's profit, taking into account randomness of the rising demand, and decrease of network element costs according to time. We then propose two decision-making methods.

In the first method, we develop an **actualization algorithm** that calculates the optimal investment time using the above described analytical modeling, as in our previous work (Morlot et al. (2007)). We base our method on the idea that the upgrade should be performed when the loss of profit, derived using analytical capacity expressions, exceeds the expected discount. As time goes by, we have realizations of the daily demand. These values can thus be used to derive a more precise optimal date than the one we found at time $t = 0$ assuming demand randomness. Instead of calculating expectations, we can now take in account what *really* happens during the period. This is the idea of the actualization algorithm, where the expected investment time is calculated using simulations.

We then introduce a second, more analytical method, based on real options to hedge against the risk that demand evolves in an unexpected way, leading to a premature investment decision or a too late one. To perform that, we model the investment as a virtual American call that gives its owner (the mobile operator) the right, but not the obligation, to buy an equipment, until a maturity date. Given the profit analytical model and the option's parameters, we propose a backward dynamic programming method to obtain the expected best investment date.

BACKGROUND

The problem of investment under uncertainty has been investigated in several domains. Longstaff & Schwartz (2001) introduced a pricing method for american options in general (this method has been approved in Protter, Clément & Lamberton (2002)). In Laughton (1991), the authors introduced a real options method to evaluate investments in oil fields. A real options framework was also proposed for evaluating investments in sustainable development projects (see Salahaldin & Granger (2005) and Salahaldin (2007)). For a detailing of real option types and correlated bibliography, see Trigeorgis (1993), and for a general economic introduction, see Dixit & Pindyck (1994).

In the telecommunications field, the applications were scarce, although real options are relevant to telecommunications in several areas: strategic evaluation, estimation and cost modelling, as indicated in Alleman (2002).

D'Halluin, Forsyth & Vetzal (2007) presented a work on a method to determine the best investment date in a wireless network. His approach was based on dynamic programming, but he didn't introduce any risk hedging method nor an option pricing. This chapter aims at showing, using the example of an **HSDPA data network**, how the investment decision can be evaluated based on a calculation of the expected profit.

THE BASIC MODEL: OPERATOR'S PROFIT AND INVESTMENT COST

The operator profit depends on the amount of data flowed by the network. However throughputs depend on a large number of parameters and are difficult to calculate. The profit thus depends on the following parameters:

- The mean traffic demand per cell (in Mbits/sec/cell). We denote it by X_t , where the time $t = 0,1 \dots T$ varies *discretely*, for example day by day. *[0, T]* stands for the considered time period. After *T*, one considers that our network is becoming obsolete.
- The customer satisfaction before and after the upgrading (S_t^1 and S_t^2)

Traffic Demand

We assume that the network is formed by circular cells of radius *R*, with a uniformly distributed demand. We have:

$X_t = \lambda_t \times \mathrm{E}[\xi]$,

where λ_t is the arrival rate per cell at the date *t* and E[ξ] is the mean size of a typical data flow. In the following we assume that E[ξ] remains stable during *[0, T]*, so that X_t is proportional to λ_t. This model is strictly equivalent to the case where the number of active users is constant, but they initiate connections more often.

To model the evolution of X_t, let us consider it as the daily sampling of a continuous stochastic process $\tilde{X}(t,W_t)_{t\in R^+}$. Usually, one monitors the 24 demands over one hour each, keeps the second or third highest, and multiply it by a given factor. As many random phenomena related to a social behavior (e.g. Salahaldin & Granger (2005), Salahaldin (2007)), we assume that $\tilde{X}(t,W_t)$ is a geometric brownian motion (see El Karoui & Gobet (2003) page 88):

$$\tilde{X}(t,W_t) = x_0 e^{\left(\alpha - \sigma^2/2\right) + \sigma W_t}, \quad t \in R^+$$

where W_t is a standard brownian motion, *α* is the trend of the demand and σ is its volatility.

Customer Satisfaction

To evaluate the customer satisfaction, let us first compute the flow throughput $\gamma_t(r)$ a user can expect at distance *r* from the center of the cell, as carried out in Bonald & Proutière (2003).

The resource of a single downlink data channel is time-shared between active users. Denote by Φ_u the fraction of time the BS transmits to user *u*, with $\sum_u \Phi_u = 1$. The data rate of user *u* is then

$C_u(r) = C(r)\Phi_u$, where $C(r)$ is the peak data rate, obtained in the absence of any other user in the cell, i.e., for $\Phi_u = 1$. When there are *x* users, the

"fair power" sharing is defined by $\Phi_u = \frac{1}{x}$. We have:

$$C(r) = \min\left(C_0, \frac{Z}{b}\cdot\frac{\Gamma(r)}{\eta + I(r)}\right),$$

where:

- C_0 is the maximum peak rate (which depends on channel bandwidth and coding efficiency)
- *Z* is the cell chip rate
- *b* is a lower bound for energy-per-bit to noise density ratio (E_b/N_0)
- $\Gamma(r)$ is the path loss between the BS and user *u*
- η is the thermal noise to received power ratio
- *I(r)* is the interference to received power ratio.

We define the cell load as (see Bonald & Proutière (2003)):

$$\rho_t = \frac{X_t}{\pi R^2}\int_0^R \frac{2\pi r dr}{C(r)}.$$

- If $\rho_t > 1$, the cell is overloaded and one can show (Bonald & Proutière (2003)) that the number of active users grows indefinitely; any individual data rate tends to zero, and the cell is saturated
- If $\rho_t < 1$, the cell is underloaded and the number of active users tends to a finite stationary regime.

We are naturally led to introduce:

$$X_{\max} = \pi R^2 \left(\int_0^R \frac{2\pi r dr}{C(r)}\right)^{-1},$$

such that $\rho_t = X_t / X_{max}$. Hence $\rho_t < 1 \Leftrightarrow X_t < X_{\max}$. It is easy to show that the average throughput of the cell is given by (Bonald & Proutière (2003)):

$$\overline{\gamma_t} = (X_{\max} - X_t)^+, \tag{2}$$

where $x^+ = \max(x, 0)$. Note that $\overline{\gamma_t} = 0$ if and only if the cell is saturated.

Now we can compute the **customer satisfaction**, which can reasonably be supposed to depend on $\overline{\gamma_t}$. Since subjective satisfactions have been shown to be more sensitive to small variations at low throughputs than at high throughputs, Enderlé and Lagrange propose in Enderlé & Lagrange (2003) to model the customer satisfaction as a negative exponential function of the throughput:

$$S_t = e^{-\beta (X_{\max} - X_t)^+}.$$

For example, β can be chosen as:

$$\beta = \log(2) \cdot \gamma_{1/2}$$

where $\gamma_{1/2}$ is the throughput value ensuring a satisfaction of 50%. Once again, note that $S_t = 0$ if and only if $X_t \geq X_{max}$ i.e., if and only if the cell is saturated.

Daily Profit

Let us recall that X_t is sampled day by day, for example during the second or the third highest peak hour. Within 24 hours, the operator transmits $\mu \min(X_t, Xmax)$ to active users, where μ is a multiplicative factor between the peak hour and the whole day. Typically, we can consider that the peak hour represents 25 % of the total daily transfer. This leads to $\mu \approx 4.3600 = 14000$ sec. Since taxation is applied to the volume of transfers and not to the duration, the gross daily profit per cell is given by:

$$\pi^{gross} = \delta\mu \min(X_t, X_{max}),$$

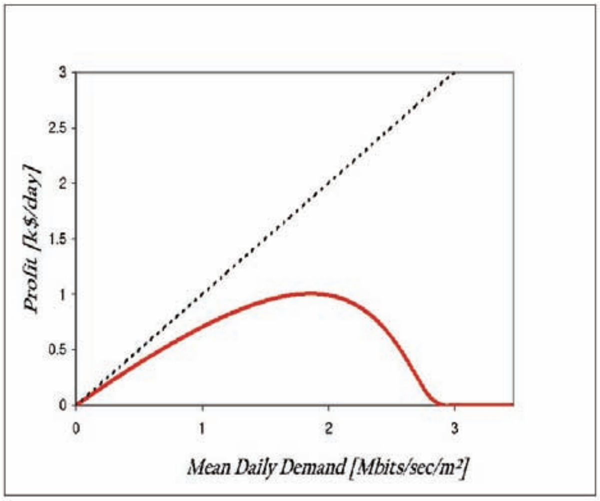
where δ is the transfer price (say in \$/Mbit). However the gross profit should be weighed by the customer satisfaction to account for the quality of the communications. The net profit is thus calculated as the product of π^{gross} by S_t:

$$\pi^{net} = \delta\mu \min(X_t, X_{\max}) e^{-\beta/(X_{\max} - X_t)^+}.$$

If $S_t = 0$, i.e. if the cell is saturated, the net profit is null. If the satisfaction is maximal, i.e., $S_t = 1$, the net profit is equal to the gross profit. To sum up we have:

$$\pi_t = \begin{cases} \delta\mu \min(X_t, X_{\max}) e^{-\beta/(X_{\max} - X_t)^+} & \text{if } X_t < X_{\max} \\ 0 & \text{otherwise} \end{cases}$$

Figure 1. Daily Profit generated by demand. The solid line represents gross profit and the dotted line represents net profit. X_{max} *= 3 Mbits/sec/cell,* δ = 0.1 \$/Mbit, μ = *14000 sec,* β *= 0.7 Mbit/sec (which corresponds to* $\gamma_{1/2} \approx$ *1 Mbit/sec)*



Intuitively, as the demand rises, X_t will increase as will the profit (Figure 1). Then, the profit will decrease because the unsatisfaction effect becomes dominant. If no upgrading action is taken, the profit will progressively tend to zero.

Once the operator decides to **upgrade**, he can install additional transmitters operating on different frequency bands. In such a case we obtain a higher value of X_{max}, so that:

$$\pi'_t = \delta\mu \min(X_t, X'_{\max}) e^{-\beta/(X'_{\max} - X_t)^+}.$$

Total Profit

If we denote the investment date by t_0 the total profit $\Pi_T(t_0)$ actualized at the date t=T is:

$$\Pi_T(t_0) = \sum_{t=0}^{t_0-1} e^{\zeta(T-t)} \pi_t + \sum_{t=t_0}^{T} e^{\zeta(T-t)} \pi'_t,$$

where ζ is the actualization rate. For simplicity, we assume that ζ is constant during the period *[0, T]*.

Upgrading Investment

The upgrading cost is a decreasing function of time. The decrease of the cost is due to many factors, for example the R&D progress, and also the serialization in the manufacturing chain. Moore's law states that electronic devices' capacity doubles every 18 months. So in this paper we assume it decreases exponentially:

$$K(t) = K_0 e^{-\varepsilon t}$$

where ε is the depreciation rate.

EVALUATING THE INVESTMENT BY AN ACTUALIZATION ALGORITHM

Total Profit

As we have now derived ρ_t analytically, we can compute numerically the expectation of $\Pi_T(t_0)$ given t_0 using the Monte Carlo method. Results for the calculation of the daily profit calculated using Scilab are shown in Figure 2.

We observe that the higher σ is , the more important the risk is and the sooner the maximal value of the expected profit is reached. Additionally, the higher σ is, the slower the curve tends to zero. When σ becomes paramount, the curve flattens out.

From the previous simulation, we can deduce the total profit shown in Figure 3. We observe that the expected profit diminishes as the volatility increases, because of the risk.

To summarize the results:

- For a given value of σ, the earlier the investment, the happier the consumers, and the higher the profit;
- At a given date, the higher σ, the lower the profit, since the system can be saturated earlier.

However, when we take the investment cost into account, the number to maximize is what we call the total Profit-minus-Cost (PMC), actualized at date *T*:

Figure 2. Expected Daily Profit during the period. Parameters are identical to those used in former figures

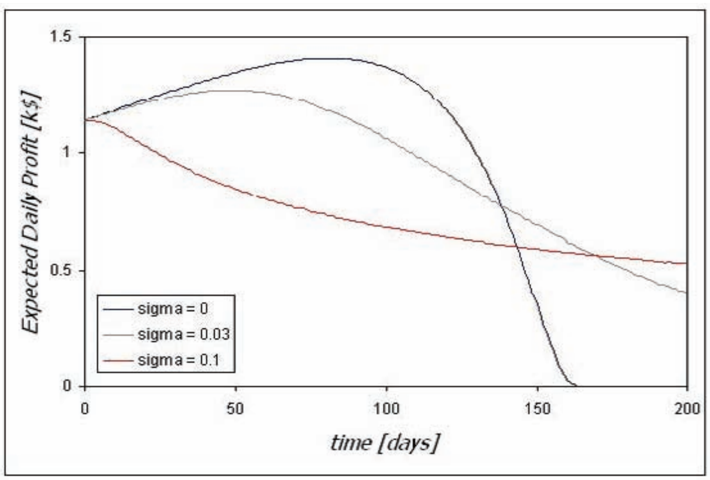


Figure 3. Total Expected Profit. X_{max} *= 8 Mbits/sec/cell after the upgrade.* $\zeta = -\log(0.95)/365\ day^{-1}$, *which corresponds to a 5% / year actualization*

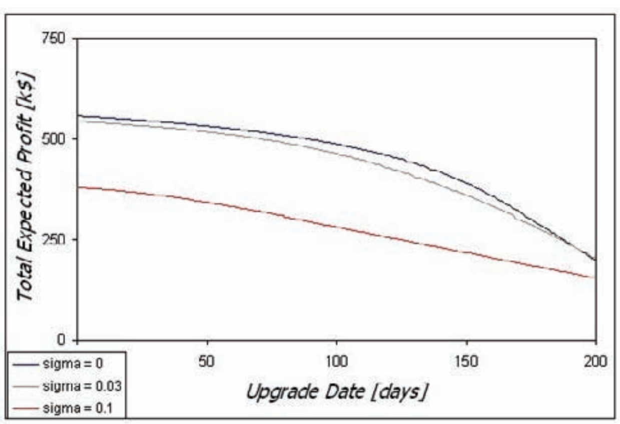
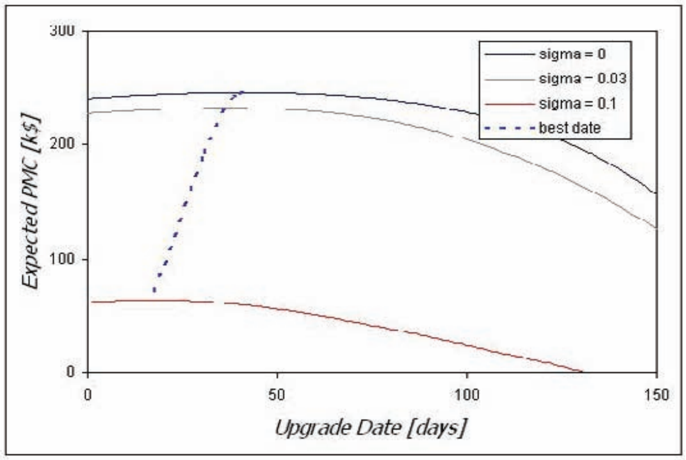


Figure 4. Expected PMC during the period. K_0 = 300 k$ and ε = -log(0.5)/365, which corresponds to a 50% / year cost decrease



$$PMC(t_0) = \Pi_T(t_0) - K_0 e^{(\zeta T - (\varepsilon + \zeta) t_0)}.$$

Note that generally ε > z, so that $\varepsilon + \zeta \approx \varepsilon$. Results when taking the *K(t)* component in our simulations and results are shown in Figure 4.

Again we show that the higher the volatility is, the lower the expected profit is. This result was expectable because of the increased risk. Furthermore, we see that the optimal investment date decreases according to the volatility, as shown in Figure 4. However, works on investment under uncertainty obtain the opposite result Salahaldin & Granger (2005). In fact, as will be shown in next section, the significant value to observe is the optimal *actualized* investment date, and we will observe the opposite trend.

Strategy

As time goes by, we have realizations of the daily demand. These values can thus be used to derive a more precise optimal date than the one we found at *t = 0* assuming demand randomness. Instead of calculating expectations, we can now take in account what *really* happens during the period.

We can implement the following algorithm: each day *t* [0, *T]*, we perform deterministic calculations over [0, t] and expectations over [t, T], with an initial value x_t of traffic demand which is

Figure 5. An example of the traffic demand evolution X_t

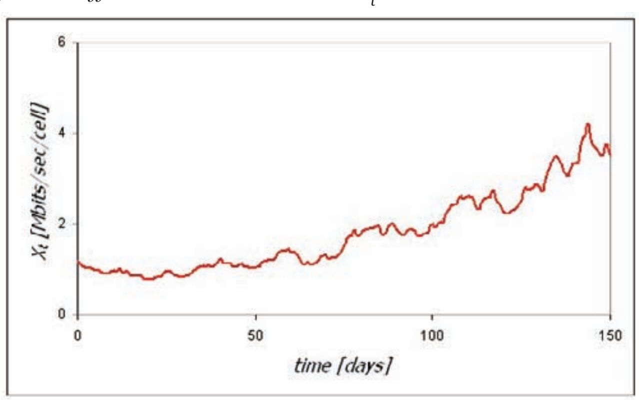
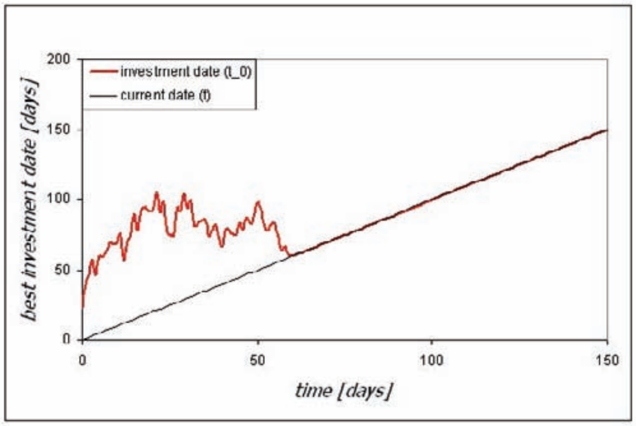


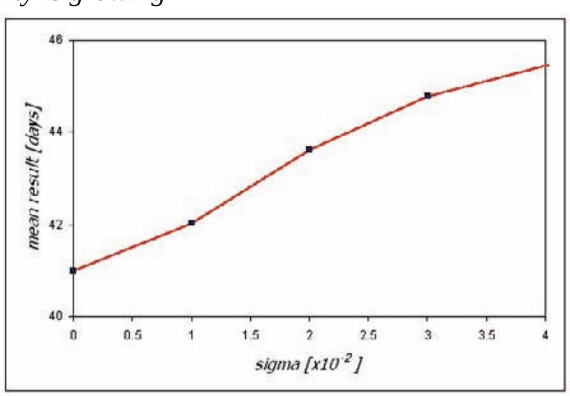
Figure 6. Best investment date during actualization algorithm in the random case (σ = 0.05 day$^{-1/2}$), for the trajectory in Fig. 5. Here, investment will take place at around t = 60 days. When X_t does not grow as fast as expected, the actualized investment date is postponed. But as soon as X_t has sufficiently grown, the actualized date is stabilized



now a deterministic parameter. Each computation thus provides us with a new investment date t_0 $[0, T]$. In other words, t_0 is daily *actualized*. When t_0 becomes close to t, it is time to invest. Note that such a situation t_0 t will inevitably happen by continuity, since at the end of the process we have $t=t_0=T$. Interestingly we can take the upgrading delay into account : if for example we need 1 month of delay for the deployment, we will plan it as soon as $t_0 - t \leq 30$ days.

Because of the randomnes, we can observe some fluctuations before t_0 becomes stuck to t (see Figure 6). To examine the impact of randomness on the investment date, we simulate several possible scenarios, in other words, several possible trajectories of the brownian demand. For each trajectory, we apply the actualization algorithm and we compute a final result, averaged over all the simulations (see Figure 7): now, we are in accordance with classic investment studies.

Figure 7. Mean final date returned by actualization algorithm regarding volatility. As expected, this date becomes later while volatility is growing



EVALUATING THE INVESTMENT BY A REAL OPTIONS METHOD

In this section, we present an alternative way to compute the optimal investment date. As shown above, there is a tradeoff between the growth of the demand (encouraging to invest) and the depreciation of the equipment cost (encouraging to wait). Then the risk is to be led to invest while the equipment is still expensive. When is the investment going to be exercised ? It depends on the additional profit expected from investing to upgrade the network: at least, this additional profit has to be greater than $K(t)$. At date t, it can be expressed as follows:

$$S_t = E\left[\int_t^T e^{-\zeta(s-t)}(\pi'(s,W_s) - \pi(s,W_s))ds \middle| F_t\right], \qquad (3)$$

which we can rewrite as $S_t = v(t,W_t)$, with:

$$v(t,x) = \int_t^T E\left[e^{-\zeta(s-t)}(\pi'(s,W_s) - \pi(s,W_s)) \middle| W_t = x\right]ds. \qquad (4)$$

Facing the decision to invest or not, the operator's strategy is to compare the profit realized if investing with the value of waiting, typically to check that the traffic is not going to decrease unexpectedly which would make the upgrading expenditure a sunk cost. This appears to be the classical problem of finding the exercise strategy for an American option, with $(S_t - K(t))^+$ as the option's payoff, denoted by $Z(t)$:

$$Z(t) = \max(S_t - K(t), 0) \qquad (5)$$

Monte-Carlo Simulations to Generate the Underlying Asset

Calculating S_t involves a complex integration (equation (3)) that cannot be performed analytically. We then use Monte-Carlo simulations as follows:

- First we compute $v(t,x)$ with equation (4) for t $[0, T]$ and x $[w_{min}, w_{max}]$. To bind efficiently the brownian motion, we use a result of El Karoui and Gobet (see El Karoui & Gobet (2003), Proposition 1.3.8.): $P(\sup_{t \leq T}|W_t| \geq c) \leq 2P(|W_t| \geq c)$, which tends to 0 extremely rapidly when c ∞. Hence, if we choose correctly our parameters so that the critic lines $W_t = w_{max}$ and $W_t = w_{min}$ lie far enough from the line $w = 0$, the probability to reach them during the experiment will be extremely low. Then, in practice, we will consider that (W_t) remains almost surely bounded.
- Then we discretize time: $t = t_0 ... t_N$ with $t_0 = 0$ and $t_N = T = N\delta t$. After that we simulate J trajectories of S_t: the j-th trajectory is denoted by S^j and has the value S_n^j at time $t_n = n\delta t$. More precisely, we simulate J trajectories of the brownian (W^j), and then we compute $S_n^j = v(t_n, W_n^j$ by interpolating $v(t,x)$. This is far more efficient than computing directly the integral, especially if we want to simulate a large number of trajectories, since we do not have to compute v each time again.

Continuation Value and Decision Tree Algorithm

At time T, the operator invests if $Z_N > 0$. More generally, at a time $t_n < T$, the operator has two alternative choices: either invest now and get Z_n, or wait and get the expected continuation value, denoted by C_n. The generated cash-flow is then given by:

$$F_n = \max(Z_n, C_n).$$

We already know Z_n by (4). As for C_n, we use the **Least Squares Monte-Carlo (LSM) approach** defined by Longstaff & Schwartz (2001).

This approach consists in writing the expected continuation value C_n as a general function of S_n (in our case we took a 2-degree polynom), taking information from the J cash-flows at t_{n+1} and using the fact that:

$$C_n(S) = e^{-\zeta \delta t} E[F_{n+1} | S_n = S],$$

where F_{n+1} is the (random) cash-flow of the option at t_{n+1}. To obtain recursively C_n, we can write the following algorithm:

- At t_N, for each trajectory $j = 1 \ldots J$, calculate the cash-flow $F_N^j = Z_N^j$.
- Move one period back to t_{N-1}. For each S^j, check if the option is "in the money", i.e. if $Z^j_{N-1} > 0$. If it is the case, calculate the continuation value C^j_{N-1} using the cash-flow if investment is delayed: $C^j_{N-1} = e^{-\zeta \delta t} F_N^j$. Estimate then the general expression of $C_{N-1}(S)$ by the LSM algorithm. This consists in regressing the found values C^j_{N-1} on a constant, S and S^2, as in Longstaff & Schwartz (2001).

Let us denote the estimated expression by $\hat{C}_{N-1}(S)$. The estimated cash-flow at *N-1* is then given by:

$$F^j_{N-1} = \max\left(Z^j_{N-1}, \hat{C}_{N-1}\left(S^j_{N-1}\right)\right).$$

If it is optimal to exerce at t_{N-1}, then by convention F_N^j becomes 0 (because the option can only be exercised once).

- For each time t_n, repeat the same process until $n = 0$.

Let us denote by O_N the set of the j such that $Z_N^j = 0$, and by I_N the set of the j such that $Z_N^j > 0$. Here is a summary of the whole algorithm:

Expected Investing Time

Finally, we obtain for each of the J trajectories a best investment date T^j_{inv}. Averaging the T^j_{inv} s, we obtain the expected investing time:

Table 1.

<table>
<tr><td colspan="2">1. simulate J trajectories (S^j)

2. for $j = 1 \ldots J$, put $F_N^j = Z_N^j$

3. for $n = N\text{-}1 \ldots 1, 0$:

for $j = 1 \ldots J$, calculate Z_n^j:
- if $Z_n^j = 0$, $j \in O_n$
- if $Z_n^j > 0$, $j \in I_n$
process O_n and I_n separately:</td></tr>
<tr><td>$\forall j \in O_n$:

put $F_n^j = e^{-\zeta \delta t} F^j_{n+1}$</td><td>$\forall j \in I_n$:

- regress $C_n^j = e^{-\zeta \delta t} F^j_{n+1}$ on *1*, S and S^2 to obtain a 2-degree polynom $\hat{C}_n(S)$
- put $F_n^j = \max\left(Z_n^j, \hat{C}_n(S)\right)$
- if $Z_n^j > \hat{C}_n(S)$, then n is the new investment date, so put $F_m^j = 0 \quad \forall m > n$</td></tr>
</table>

$$E[T_{inv}] \approx \frac{1}{N} \sum_{j=1}^{J} T_{inv}^{j}. \quad (6)$$

Numerical Results

In order to illustrate our algorithm, we applied it using the free simulator Scilab (see Gomez (1999)). We considered a **HSDPA pure data network** with a random growing demand, as described in section *Traffic Demand*. We used the following parameters for our computation:

- The investment can take place until $T = 150$ days.
- The equipment's price decreases with a rate ε of 50% per year.
- The actualization rate ζ is fixed to 5% per year.
- The traffic demand starts at $x_0 = 1.2$ Mbit / sec / cell, and increases with a drift fixed to $\alpha = 0.54\%$ per day. Its volatility varies between 0.005 and 0.02 $day^{-1/2}$. Its maximal value is fixed to X_{max} =3 Mbit/sec/cell before the investment, and to X'_{max} =8 Mbit/sec/cell after the investment.
- The data transfer price is fixed to $\delta = 0.1$ \$/Mbit.
- We take a satisfaction parameter of $\beta = 0.7$ Mbit/sec/cell.
- We simulate 5000 different trajectories of the asset.

On Figure 8, we represent the investment date versus K_0. Recall that the date is obtained with equation (6), where K_0 implicitely appears in generalized function Z (see equation (5)). It appears that the investment date is very low for lower values of K_0. This happens because the equipment is already quite cheap at the beginning, thus it is all the more interesting to invest early.

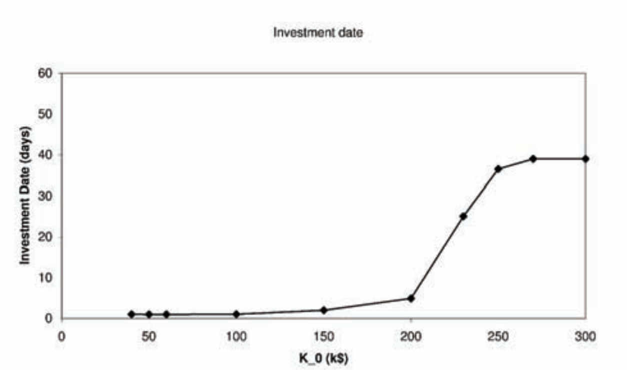
Of course, the higher K_0 is, the later the investment takes place, since one has to wait longer until the equipment becomes affordable.

COMPARISON OF BOTH METHODS

In this section we compare results obtained with the actualization algorithm on the one hand, and the **backward programming** method on the other hand, with $K_0 = 300$ k\$ and $\sigma = 0.01$ $day^{-1/2}$.

- With the former method, we find an expected investment date of 42.04 days with a standard deviation of about 0.1 day.
- With the latter, we find an expected investment date of 39 days with a standard deviation of about 1.15 day.

Figure 8. Investment date



The results are very close, and thus, both methods can be used for estimating the investment date. However, the real option method has the advantage of being quicker in terms of computation duration.

FUTURE TRENDS

The above analysis shows that uncertainty is an important issue while dealing with infrastructure investment in telecommunication networks. In fact, it has been shown that waiting is the optimal choice in several cases. The methods presented herein should be of interest for operators, allowing them to make optimal decisions as to when upgrade their radio networks.

So far, the decision methods have been illustrated in a particular situation. However it could be applied to a wide range of different situations.

- First of all, we could generalize the analytical model proposed in section "Basic Model". For example, we could introduce:
  - Admission control: if the network guarantees to each user a minimal data throughput, some users may be blocked. In that case, the satisfaction S_t depends on the blocking rate instead of the throughput. Bonald & Proutière (2003) gives the blocking rate by:

    $$B = \frac{\rho_t^m}{1 + \rho_t + \ldots + \rho_t^m},$$

    where rt is the cell load and m is the maximum allowed number of users in one cell. Then we would assume that the customer satisfaction decreases exponentially according to blocking rate.
  - Different power allocations. Then we would have to balance satisfaction over the cell by the probability density function of the user position.
  - Discontinuities and randomness in the investment cost curve. So far, we have supposed this latter curve being continuous and deterministic, but due to R&D's breakthroughs, it may not be the case. Of course, we would have to adapt the PMC's computation.
- Furthermore, we could imagine the operator planning more than one investment (e.g. adding WiFi hotspots): a multi-investments model would have to be deployed. Basically, this could consist in introducing two investment dates t_1 and t_2, corresponding to two technologies characterized by X'_{max} and X''_{max}, with $X_{max} < X'_{max} < X''_{max}$. Both methods may then be turned into 2-dimensional versions.
- A last generalization consists in effectively buying an option from a financial institution, that gives us the right to buy the equipment at a fixed price, possibly lower than the real market price. The backward analysis can be extended to price this option, for example with a risk-neutral approach.

CONCLUSION

In this chapter, we developed two methods to evaluate upgrading investment date in a data network. In the first method, we considered that this date is the result of a compromise between the decrease of upgrade investment cost with time and the loss of profit generated by insufficient capacity. The upgrade should hence be performed when the loss of profit, derived using analytical capacity ex- pressions, exceeds the expected discount. We then developed an actualization algorithm that calculates the optimal investment time using the above described analytical modeling.

The second method is based on a real options approach as the operator has the right, but not the obligation, to perform the upgrading investment, under the demand (traffic) uncertainty. We showed how the problem can be formulated as a real options problem. In fact, facing the decision to invest or not, the operators strategy is to compare the profit realized if investing with the value of waiting, typically to check that the traffic is not going to decrease unexpectedly which would make the upgrading expenditure a sunk cost. This appears to be the classical problem of finding the exercise strategy for an American option. We thus developed a dynamic programming algorithms that calculates the optimal investment date and the expected operator's profit. We finally compare the results obtained from both methods and show the advantages and drawbacks of each of them.

ACKNOWLEDGMENT

The work was supported by ANR project Winem.

REFERENCES

Alleman, J. (2002). *A New View of Telecommunications Economics*. Telecommunications Policy.

Bonald, T., & Proutière, A. (2003). Wireless Downlink Data Channels: User Performance and Cell Dimensioning. *Proceedings of ACM MobiCom'03*, September 14–19, 2003, San Diego, California, USA.

D'Halluin, Y., Forsyth, P. A., & Vetzal, K. R. (2007). Wireless Network Capacity Investment. *European Journal on Operational Research, 176*(2007), 584-609.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

El Karoui, N., & Gobet, E. (2003). *Introduction au Calcul stochastique, première partie*, Imprimerie de l'Ecole Polytechnique.

Enderlé, N., & Lagrange, X. (2003). User Satisfaction Models and Scheduling Algorithms for Packet-Switched Services in UMTS. *Proceedings of IEEE VTC spring*, April 2003.

Gomez, C. (1999). *Engineering and Scientific Computing with Scilab*. ISBN: 978-0-8176-4009-5, 1999.

Laughton, D. (1991). The Valuation of Offshore Oilfield Development In Leases. Advances In M. Breton &G. Zaccour (Eds.),Operations Research in the Oil and Gas Industry and Editions Technip (pp.63-82), Paris 1991.

Longstaff, F. A., & Schwartz, E. S. (2001). Valuing American Options by Simulation: A Simple Least-Squares Approach. *The Review of Financial Studies, 14*(1), Spring.

Morlot, F., Fourestié B., & Elayoubi, S. E. (2007). Optimizing the date of an upgrading investment in a data network. *Proceedings of the IEEE Vehicular Technology Conference IEEE VTC-fall 2007*, Baltimore, October 2007.

Protter, P., Clément, E., & Lamberton, D. (2002). An analysis of a least squares regression method for American option pricing. *Finance and Stochastics, Springer, 6*(4), 449-471.

Salahaldin, L. (2007). A Numerical Algorithm for Decision-making in Sustainable Transport Projects Investment. *Proceedings of the 11th Annual International Conference on Real options*, Berkeley, CA, June 2007.

Salahaldin, L., & Granger, T. (2005). Investing in Sustainable Transport to Relieve Air Pollution under Population-Growth Uncertainty. *Proceedings of the 9th Annual International Conference on Real options*, Paris, June 2005.

Trigeorgis, L. (1993). The Nature of Options Interactions and the Valuation of Investments with Multiple Real Options. *Journal of Financial and Quantitative Analysis, 28*(1), 1-20.

KEY TERMS

Actualization Algorithm: This is the first method presented above. If an optimal investment date is found at a given time, then it is wise to actualize it regularly in the future, taking into account new information about traffic evolution.

Dynamic Programming: After simulating several trajectories of the traffic demand, dynamic programming consists in building a decision tree in a time-backward manner.

Operator Gross Profit: This is the cash the operator can expect to charge his users for downloading a given amount a data.

Operator Net Profit: This is equal to gross profit ponderated by user satisfaction.

PMC: This means "Profit minus Cost". When the operator considers his total profit over a given period, taking into account a necessary investment, he has to deduce the equipment cost.

Real Option: It gives the operator the choice between investing at once, or waiting for hopefully better conditions.

Upgrade Investment: If the user satisfaction becomes low, the operator net profit decreases, and the operator has to plan an investment, in order to increase the capacity of the network.

User Satisfaction: This is an exponential ponderation coefficient of the operator gross profit. It takes into account the user being frustrated if his throughput is too low.

Chapter XXVI
Telecommunication Investments Analysis: A Multi-Criteria Model

Georgios N. Angelou
University of Macedonia, Greece

Anastasios A. Economides
University of Macedonia, Greece

ABSTRACT

Recognizing the inadequacy of traditional quantitative cost-benefits analysis for evaluating telecommunications investments, researchers suggest real options (ROs) for controlling and valuating telecommunications business activities. However, ROs are based on the concept to delay investment for collecting more information and learning more about business conditions and during this delay another competitor may act gaining significant competition advantage. In addition, RO models are strictly quantitative and very often telecommunications investments may contain qualitative factors, which cannot be quantified in monetary terms. In addition, ROs analysis results in some factors that can be treated more efficiently when taken qualitatively. This work deals with quantitative and qualitative analysis and integrates ROs and Analytic Hierarchy Process (AHP) into a common decision analysis framework providing a multi-criteria model, for analyzing telecommunication investments in the deregulated business field.

INTRODUCTION

Telecommunications markets all over the world recently have been, and still are, undergoing drastic changes, fuelled my market reforms and technological progress. State-owned monopolists have been privatized and markets have been liberalized. These transformed markets have attracted entrants in many varieties. Some entrants roll out complete networks, while other build only partial networks or perhaps offer services without having infrastructure themselves but having access to the networks of incumbent operators.

The valuation of telecommunication business activities is a challenging task because it is characterized by high-level uncertainty. In addition, these business activities are not possessed exclusively by a single firm but rather are shared by many competitors. The main challenge for a potential provider (investor) is to roll out its business activity at the right time and the right scale. The time and scale depend on telecommunication services penetration, network infrastructure cost, area characteristics, applications offered, expected tariff evolution, customers' willingness to pay, demand forecasts, evolution of expected market shares and investor's technical skills.

Although the traditional cost-benefit analysis expressed in money terms such as investment cost, revenues, and Net Present Value (NPV) are significant factors to be taken into account in the analysis process, it is by no means sufficient for capturing the complexity of the problem in its entire. It is not able to accommodate the flexibility, for example, to defer and to abandon an investment plan at certain discrete pre-specified points in time.

Real Options (ROs) address this inadequacy and offers management the flexibility to take actions, which can change aspects of the business activity over time. ROs have been proposed for business analysis and risk management (Trigeorgis, 1996).

ROs applications to risk management and investment evaluation of telecommunication business field have mainly focused on a single and a-priori known option. However, these options are not inherent in any telecommunication investment. Actually, they must be carefully planned and intentionally embedded in the telecommunication investments in order to mitigate its risks and increase its return. Moreover, when a telecommunication investment involves multiple risks, by adopting different series of cascading options we may achieve risk mitigation and enhance investment performance. Given the investment's requirements, assumptions and risks, the goal is to maximize the investment's value by identifying a good way to structure it using carefully chosen real options (Benaroch, 2002).

However, ROs application in telecommunication business field raise some issues that require attention. ROs models are strictly quantitative, while telecommunication analysis may also involve qualitative factors. In addition, ROs analysis itself brings to the "surface" a number of factors that cannot be quantified, at least easily, by existing ROs models and methodologies.

In this work we discuss all these issues and include them in a multi-criteria perspective providing specific decision analysis models. We consider that a holistic methodology is required in order to assist executives and decision makers in formulating problem parameters, understanding their interactions, estimating their contribution to the overall business value and so valuating effectively new telecommunication business activities.

We discuss limitations of the quantitative analysis, in the specific field proposing qualitative factors that can be combined in a multi-criteria analysis. For integrating quantitative and qualitative factors in one utility function we adopt Analytic Hierarchy Process (AHP).

The proposed analysis provides a better understanding of the telecommunication risks and the various qualitative factors inherent in such investments, enabling these investments to be deployed more optimum and valued with higher accuracy. Our main contributions are the following:

- We integrate financial quantitative, qualitative and risk factors as provided by the ROs and competition threat in an Analytic Hierarchy Process (AHP) structure.
- We provide a risk management framework for telecommunication investments based on ROs thinking.

Our aim is to find the business deployment that achieves a balance between risk control and performance maximization under competition.

The rest of the paper is organized as follows. In Section 2, we provide background for the involved techniques. In Section 3, we provide the proposed methodology and model. In Section 4, we apply the proposed decision analysis framework to a real life case study. In Sections 5 and 6 we conclude and suggest possible future research.

BACKGROUND

ROs Background Material

An option gives its holder the right, but not the obligation, to buy (call option) or sell (put option) an underlying asset in the future. Financial options are options on financial assets (e.g. an option to buy 100 shares of Nokia at the pre-agreed price 90€ per share on January 2008). ROs approach is the extension of the options concept to real assets (Trigeorgis, 1996). For example, a telecommunication investment can be viewed as an option to exchange the cost of the specific investment for the benefits resulting from this investment. An investment embeds a RO when it offers to the management the opportunity to take some future action (such as abandoning, deferring or expanding the investment) in response to events occurring within the firm and its business environment (Trigeorgis, 1996). Telecommunication research on ROs recognizes that such investments can embed various types of ROs, including: defer, stage, explore, alter operating scale, abandon, lease, outsource, and growth. Trigeorgis (1996) provided an in-depth review and examples on different real options. For more practical issues the reader is referred to Mun (2002). Especially, for the telecommunication business field Iatropoulos et. al. (2004) applied ROs analysis for analyzing broadband technology investments. Finally, Angelou and Economides (2005) provided an extended survey of ROs applications in real life telecommunication investments.

Real Options Limitations and Need for Qualitative Perspective Integration

In business practice, several conceptual and practical issues emerge when trying to use options theory as proposed in the current information communication technology (ICT) literature. It is accepted that all ROs models provide approximate valuations of ROs values (Renkema, 1999).

Even the so-called accurate ROs models such as the Black-Scholes formula require some assumptions whose validity is still under criticism in the field of telecommunication investments. In particular, while ROs analysis is widely proposed for evaluating such investments, it is still accepted that ROs applicability is limited by the fact that ICT investments assets are not traded. The non-tradability of ICT assets cannot reveal the investor's risk attitudes to estimate the correct discount factor of ICT investments. The theoretical foundation of the ROs analysis and its relevance to ICT investments has been discussed and applied in practice by Benaroch and Kauffman (1999, 2000) as far as the real asset non-tradability issue and risk-neutrality of the investor are concerned. However, its limitation is still under discussion.

Existing models for ROs valuation assume a certain distribution of the resulting cash flows, based on an efficient market. However, this is rarely the case in the context of investments in the ICT business field, which is known for its uncertain and unpredictable conditions. It has also been recognized that finance-oriented option valuation models are too complex for managerial decision-making practice, when real life business conditions are considered Fichman et. al. (2005). In particular, after the liberalization of the telecommunication market, the required competition modeling has increased the complexity of existing options models. It is very difficult for senior managers to accurately estimate the parameters of a statistical distribution of outcomes and mainly volatility since they do not really have a "gut feel" for the estimation of the volatility, even though they understand its technical definition as a statistic. Options theory in its present state does provide a conceptual decision framework to evaluate technology investments but, in many cases, cannot be considered as a fully operational tool for management. If it is expected that practitioners and senior managers will resist the use of formal options pricing models, then the qualitative option valuation can be an alternative analysis process. This is based more on the intuition of decision makers and forecasting for risks profiles, and less on sharply quantified prediction for parameters used in the formal options models.

Overall, these issues suggest that even quantified ROs analysis could produce only approximate valuations, which in some cases can cause serious mistakes in investment decisions (Fichman et. al., 2005). For these reasons, we may adopt typical DCF techniques such as NPV instead of ENPV value and combine tangible factors with qualitative ROs thinking. Hence, our multi-criteria decision analysis framework can be extended including typical tangible factors from financial perspective and intangible ones from qualitative ROs thinking.

A Brief AHP presentation and Literature Review

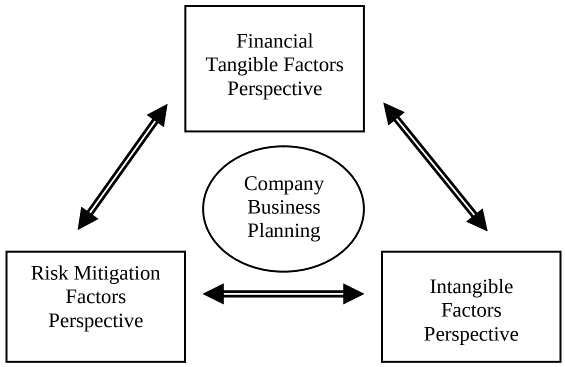
AHP is a multi-criteria decision analysis technique. It aims at choosing from a number of alternatives based on how well these alternatives rate against a chosen set of qualitative as well as quantitative criteria Saaty and Vargas (1994), (Schniederjans,2005). Using AHP, it is possible to structure the decision problem into a hierarchy that reflects the values, goals, objectives, and desires of the decision-makers. Thus, AHP fits the strategic investments problems and the framework of this study. The main advantage of the AHP approach is that different criteria with different measures can be easily transformed into a single utility measure. As input, AHP uses the judgments of the decision makers about the alternatives, evaluation criteria, relationships between the criteria (importance), and the relationships between the alternatives (preference). In the evaluation process, subjective values, personal knowledge, and objective information can be linked together. As an output, the goal hierarchy, the priorities of alternatives and their sensitivities are reached. Concerning, examples of AHP application in information, communication technology literature, Bodin et al. (2005) proposed the AHP method to determine the optimal allocation of a budget for maintaining and enhancing the security of an organization's information system. Hallikainen et al. (2002) proposed an AHP-based framework for the evaluation of strategic IT investments. They applied the principles of AHP to compare a number of Information Technology investment alternatives. Tam and Tummala (2002) formulated and applied an AHP-based model for selecting a vendor for a telecommunications system. Lai et al. (1999) applied the AHP to the selection of a multimedia authoring system. Finally, Kim (1998) used the AHP to measure the relative importance of Intranet functions for a virtual organization.

THE PROPOSED MODEL-METHODOLOGY

Next, we present a methodology that helps to address the question: *How can we control firm, market and competition risks so as to configure a specific telecommunication business activity in a way to minimize risk and increase performance?*

The proposed model contains three perspectives, financial tangible factors (FTF) perspective, risk mitigation factors (RMF) perspective, and intangible factors (IF) perspective (Figure 1).

Figure 1. The proposed model – three perspectives



Financial Tangible Factors Perspective

The financial perspective evaluates how a company is meeting though financial measures. For the valuation of financial perspective we may adopt traditional accounting techniques such as Net Present Value (NPV), Return on Investment (ROI) and Internal Rate of Return (IIR). We also include the Expanded Net Present Value (ENPV), which is the business value as estimated by ROs analysis. As mentioned before in the deregulated business field typical ROs models do not sufficiently model the competition threat. We propose the model of Angelou and Economides (2007B) for estimating the business value with ROs under competitive conditions in a business environment with may players. They adopt exogenous competition modeling considering that competitors' arrival rate and competition impact from each entry into the market follow a joint-diffusion process with business revenues.

Intangible Factors (IF) Perspective

Intangible factors are difficult, if not impossible, to quantify in absolute monetary terms, but are still important to the decision making process. Particularly, ROs analysis itself brings to the "surface" a number of factors that cannot be quantified, at least easily, by existing ROs models and methodologies. Fichman et al. (2005) called them potential pitfalls of option thinking for risk management and investment evaluation. We integrated some of them, in our analysis, in order to achieve a balance between risk control achieved by options adoption and other issues influencing the overall investment's deployment strategy and limit the options thinking applicability.

Among others, not all investments can be divided into stages implementing stage and expand options. Sometimes a firm should consider an investment as a whole entity, such as when external funds must be raised or when co-investment from other parties is required. Another issue is that stakeholders may prefer all at once funding to obtain maximum control of the investment and have so more time to get a troubled investment back on track before facing a next track of justification. We introduce this possibility in our analysis by considering the intangible factor "*Capability-Interest of staging the investment*" (CSI).

In ROs literature investment opportunities, known in advance, based on initial infrastructure investments are treated as growth options, while for the estimation of their values compound option models are utilized. However, telecommunication growth investment opportunities in reality can be hardly defined during decision phase (Benaroch, 2002). For this reason, we model qualitatively the existence of growth investment opportunities, which are based on investments in previous phases of a firm's business activity and cannot be defined quantitatively in advance.

Concerning growth options, the main challenge is the difficulty of estimating their values (due to ambiguity of future cash flows) and uncertainty about the appropriate value for option model parameters. We name this intangible option factor as "*no clarified growth options*" (NCO). Also, building in option to abandon or contract operation may concern intangible costs related to credibility and morale. We model this possibility by the intangible factor "*cost of scaling down operation*" (CSO). Finally, a potential pitfall of switch-use option is that it can add extra time and expense to the development of the initial information communication technology platform in order to change from shadow to real option. Creating this option (making it real) usually involves making the ICT platform more generic and modular for obtaining higher flexibility, experiencing however higher cost. We model this issue as intangible factor named "*cost of systems flexibility-modularity*" (CSF).

Another factor that can be integrated in a future work is the higher uncertainty clearness-control (UC) during waiting period. In our model, we consider the amount and type of uncertainty control achieved by each of the portfolio's projects. We do not want to substitute the UC, achieved by the ROs analysis and quantified by the volatility of the stochastic parameters, such as investment revenues V and one time investment cost C (σv, σc). However, the overall uncertainty of an investment opportunity cannot be easily quantified. For example, the uncertainty of customers' demand may be quantified by estimating its contribution in the overall investment's volatility, while the contribution of technology and the firm's uncertain capability to optimally exploit investment benefits may not. By adopting qualitative analysis, we can model some of the uncertainties inherent in the investment opportunity that cannot be quantitatively estimated and included in the overall investment's volatility.

Benaroch (2002) provided a method for estimating the overall investment's uncertainty (volatility), which can be broken down into its components (e.g., customers' demand uncertainty, competition's uncertainty and technology's uncertainty). However, the estimation of each component of the uncertainty may be impossible. We may extend this work by considering that some of the overall components of the uncertainty may be treated as qualitative factors, while the sources of uncertainty that can be quantified and included in the estimation of the overall volatility can be integrated into the typical ROs models. Angelou and Economides (2008) provide an extensive discussion of these subjects.

Risk Mitigation Factors Perspective

Risk management strategies are oriented towards identifying different types of risks, assessing their relative importance for the project, and implementing strategies for managing risks (Kumar, 2002). Risk management actions can be viewed as being of two types. The first is oriented towards reducing the degree of risk; for example, a major source of uncertainty in IT projects is uncertainty regarding the scope or specifications of the project. This can be partially resolved by interviewing multiple stakeholders. However, since risk cannot be completely eliminated, a second type of strategy, oriented towards hedging risks is important. Risk hedging strategies are insurance-like ones oriented towards minimizing the negative impact of risk, when the associated uncertainty is resolved over time. For example, specification uncertainty in IT projects may be due to uncertain business conditions that may be resolved as the project progresses (Kumar, 2002). Telecommunication risks can be placed into three categories (Benaroch, 2002; Bräutigam and Esche, 2002).

- Firm-specific risks are due to uncertain endogenous factors (endogenous or technical uncertainty). They could be the result of uncertainty about the ability of the firm to fully fund a long-term capital-intensive investment, the adequacy of the firm's development capabilities to a target investment, the fit of the target application with various organizational units, etc. These factors affect the ability of the investing firm to successfully realize an investment opportunity.
- Competition risks are the result of uncertainty about whether a competitor will make a preemptive move, or simply copy the

investment and improve on it. These risks give rise to the possibility that the investing firm might loss part or all of the investment opportunity.

- Market risks are due to uncertain exogenous factors that affect every firm considering the same investment (exogenous or market-related uncertainty). These risks could be the result of uncertainty about customer demand and prices for the products or services, a target investment yields, potential regulatory changes, unproven capabilities of a target technology, the emergence of a cheaper or superior substitute technology, and so on. These factors can affect the ability of the investing firm to obtain the payoffs expected from a realized investment opportunity.

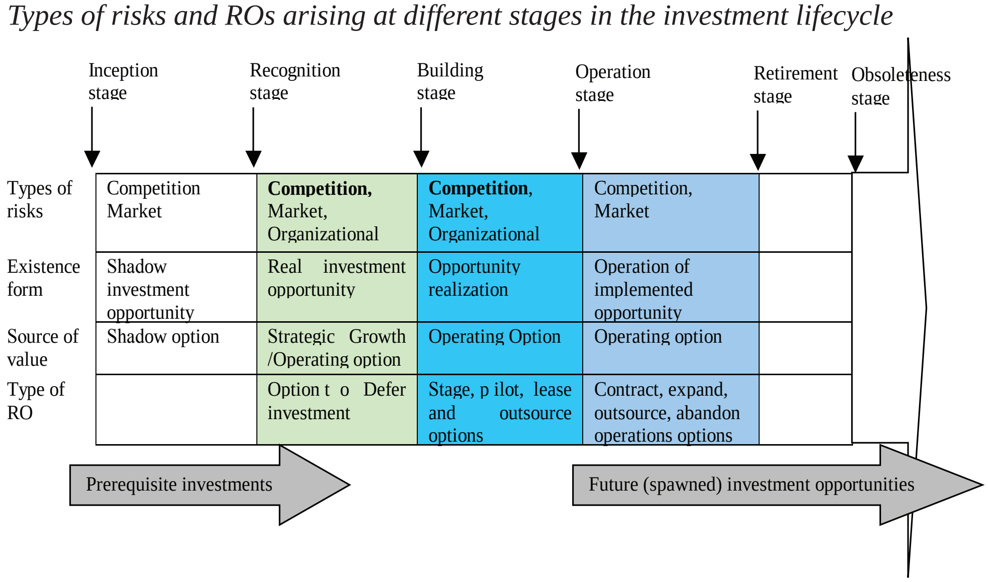
Research on technology investment's evaluation and risk management recognizes that ROs thinking emphasizes the sources of risks inherent in such investments and contributes to risk control.

Life-Cycle of Investment Opportunity Real Options

The lifecycle of an investment starts at the inception stage. During this period the investment exists as an implicit opportunity for the firm that can be facilitated by a prerequisite investment (Figure 2). The firm possesses a shadow option. During the recognition stage, call "Wait-and-See" (WaS) period, the investment is seen to be a viable opportunity. The opportunity can be treated as a RO. The building stage follows upon a decision to undertake the investment opportunity. In the operation stage, the investment produces direct and measurable payoffs. Upon retirement, the investment continues to produce indirect payoffs, in the form of spawned investment opportunities that build on the technological assets and capabilities it has yielded. When these assets and capabilities no longer can be reused, the investment reaches the obsoleteness stage. Each stage of the investment opportunity is relevant to a number of operating and growth ROs, such as option to defer, stage, lease, expand (Benaroch, 2002). The reason is that each type of RO essentially enables the deployment of specific responses to threats and/or enhancement steps. In addition, each stage of the investment is also experiencing a variety of risks. For controlling these risks we can adopt a number of investment modes

Defer investment to learn about risk in the investment recognition stage. If we don't know how serious some risk is, the option to *defer* investment permits learning about the risk by acquiring

Figure 2*. Types of risks and ROs arising at different stages in the investment lifecycle*



information passively (observe competitor moves, review emerging ITs, monitor regulatory actions, etc.) or actively (conduct market surveys, lobby for regulatory changes, etc.). Such *learning-by-waiting* helps to resolve market risk, competition risk, and organizational risk. Apparently, the greater the risk, the more learning can take place, and the more valuable is the deferral option.

Partial investment with active risk exploration in the building stage. If we don't know how serious some risk is, investing on a smaller scale permits to actively explore it. Three options facilitate *learning-by-doing*, that is, enable gathering information about the firm's technological and organizational ability to realize the investment successfully. The option to *stage* investment supports learning via a sequential development effort, and the options to *pilot* and *prototype* support learning through the production of a scaled down operational investment. The last two options compress the investment lifecycle, thus allowing to learn early how competitors, customers, regulatory bodies and internal parties will react to the investment initiative. Put another way, these options permit market risk, development risk and organizational risk to be transferred to earlier parts of the full scaled investment lifecycle. Similarly, the *stage* option divides the investment realization effort into parts, thus permitting to transfer risk across parts within the building stage. For example, implementing the riskiest parts of the realization effort as early as possible helps to reveal up-front whether the entire realization effort can be completed successfully (e.g., within schedule and budget).

Full investment with reduction of the expected monetary impact of risk in the building and operation stages. Here, options help to lower the value consequences of risk and/or the probability of its occurrence. An example of the former is the option to *lease* development resources, which protects against development and market risks by allowing to kill an investment in midstream and save the residual cost of investment resources. A way to lower the probability of risk occurrence is the option to *outsource* development. This option lowers the risk of development failure by subcontracting (part or all of) the realization effort to a third party that has the necessary development capabilities and experience. In essence, both these options permit *transferring risk* (partially or fully) to a third party.

Dis-investment/Re-investment with risk avoidance in the operation stage. If we accept the fact that some risk cannot be actively controlled, two options offer contingency plans for the case it will occur. The option to *abandon* operations allows redirecting resources if competition, market or organizational risks materialize. The option to *alter scale* allows contracting (partially disinvest) or expanding (reinvest) the operational investment in response to unfolding market and organizational uncertainties.

Based on the logic of these investment modes, the mapping of specific risks to specific options that control them can be refined to fit any class of IT investments. Examples of ROs thinking in the basis of the aforementioned investment modes are provided by Benaroch (2002) and Angelou and Economides (2007 A). Particularly, they apply ROs thinking based on the aforementioned investment modes for analyzing information technology investment opportunities. ROs analysis can control different sources of risks existing in the various stages of the investment life-cycle. We use a classification for analytical definition of telecommunication risks based on Benaroch (2002) and (Bräutigam & Esche 2002) proposals. Table 1 shows the main sources of telecommunication risks as well as their mapping to the specific ROs that can control them.

In the following we propose an AHP structure in order to combine all the aforementioned factors into one utility function.

Table 1. Risk factors inherent in telecommunication investments and options that can control them

Risk Opportunity	Recognition	Building				Operation			
	Defer	Stage	Explore/Pilot	Outsource Development	Lease	Abandon	Contract	Expand	Outsource
F1 firm cannot afford the project (unacceptable financial exposure)	+		+	+	+				
F2 costs may not remain in line with projected benefits	+		+	+	+	+	+		
P1 staff lacks needed technical skills	+	+	+	+	+				
P2 project is too large or too complex		+	+	+	+				
P3 inadequate infrastructure for implementation		+	+			+	+		
P4 the project is not on Time		+	+		+	+			
F1 wrong design (eg, analysis failed to assess correct requirements)		+	+	+		+			
F2 problematic requirements (stability, completeness, etc.)		+	+			+			
O1 uncooperative internal parties	+	+	+			+			
O2 parties slow to adopt the application		+	+			+	+		
C1 competition's response eliminates the firm's advantage	+	+	+		+	+	+		
C2 competition acts before the firm	+		+						
E1 low customer demand, with inability to pull out of market	+	+	+	+	+	+	+		
E2 demand exceeds expectations (follow-up opportunities exist)	+		+					+	
E3 too high customer response may overwhelm the application	+		+		+	+	+	+	
E4 customers may (bypass) develop their own solutions	+		+			+	+		
E5 unanticipated action of regulatory bodies	+				+	+			
E6 Price uncertainty	+					+			
E7 environment changed requirements (expected benefits vanish)	+		+			+			
E8 Other factors such as Legal issues, Natural Phenomena, Social issues, Armed conflicts, Taxation.	+					+			
T1 application may be infeasible with the technologies considered		+	+	+	+	+			
T2 the introduction of a new superior implementation technology may render the application obsolete	+				+	+			
T3 the implementation technologies considered may be immature	+		+	+		+			

INTEGRATING QUALITATIVE REAL OPTIONS WITH AHP MODEL

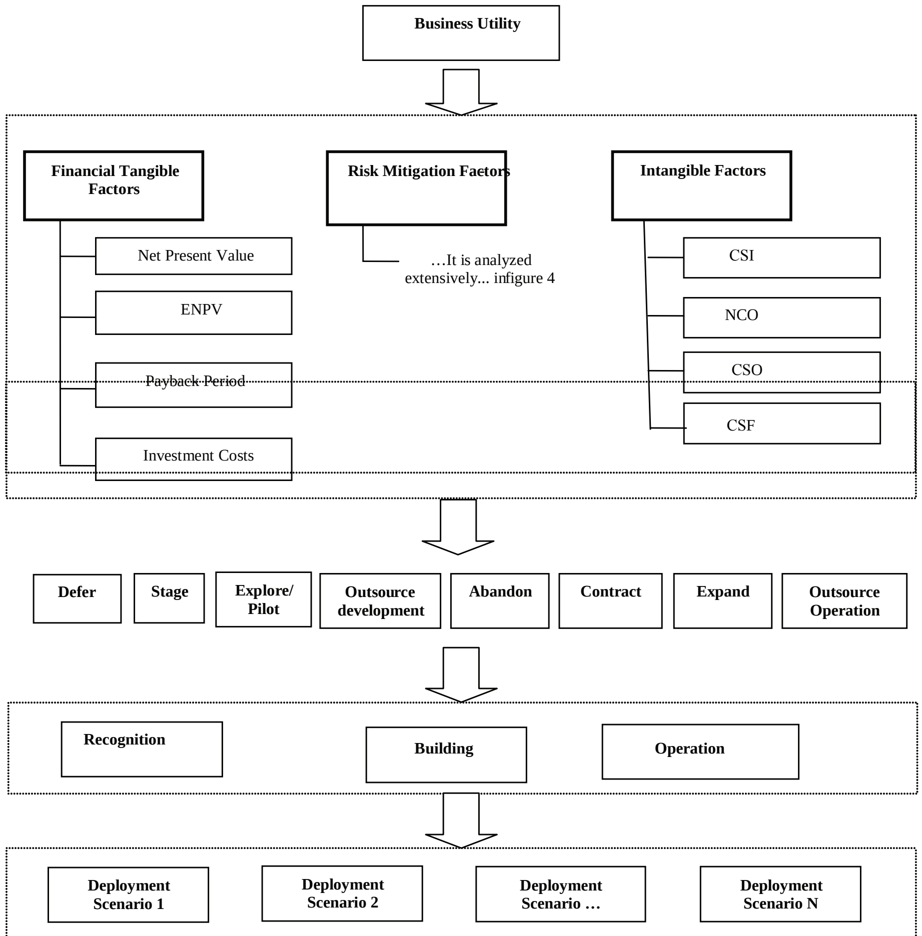
The structure of the decision analysis framework contains four levels: i) the content of the specific investment opportunity, which can be deployed in various ways, ii) the life-cycle stages of the investment opportunity, iii) the options level that is embedded in each of these stages and mapped to specific types of risks, iv) the multi-criteria level that contains financial tangible, risk and intangible analysis (Figure 3). The overall utility of AHP structure is composed by these criteria (factors), which may be further decomposed into their applicable sub-criteria. We apply the pair-wise comparisons for each of these sub-criteria. The final result of the analysis, at the top, is the prioritization of the various deployment scenarios according to the overall firm business utility.

The analytical view of the Risk Mitigation sub-module is deeply analyzed in Figure 4. The criteria used in our structure are coming from Table 1 and indicate the risk inherent in telecommunication investments. Analytically, we perform pair-wise comparisons of the deployment scenarios for each of the risk factors focusing on the risk control that each scenario can provide. The pair-wise comparisons concern the amount of risk that is resolved and controlled, depending on the option(s) existence in each scenario. Our target is to select the deployment scenario that provides the highest value for the risk mitigation utility.

Summary

The proposed methodology involves four main steps that must be repeated over time as more information is collected concerning the overall business environment. These steps help to opti-

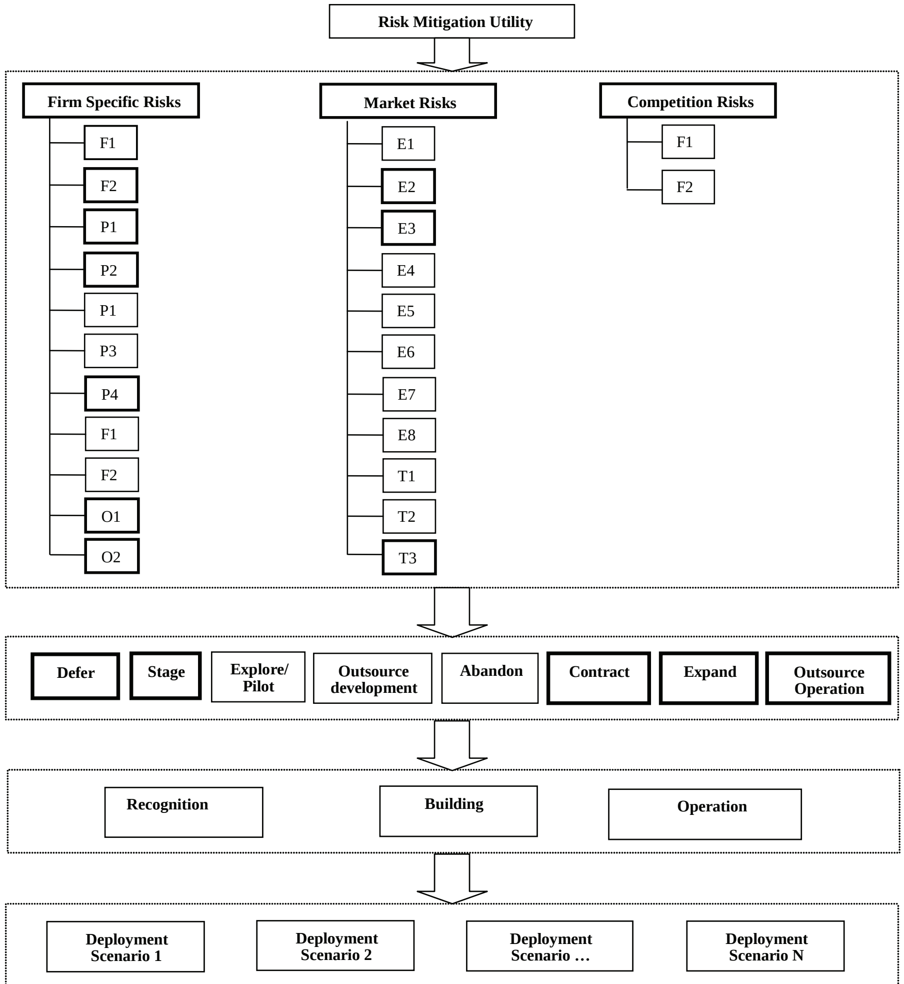
Figure 3. Analytical view of the decision analysis framework



mally configure the investment under the information set available initially, but as time passes they must be re-applied in case that some risks get resolved or new risks surface. In the following we present these steps.

- Define the content of the overall business activity and its risk profile. State the investment goals, requirements and assumptions (technological, organizational, economic, etc.), and then identify the risks inherent in the investment.
- Recognize the options mapped to specific risks and use them to adopt investment modes to be examined.
- Evaluate investment-structuring alternatives and find a subset of the recognized options that maximally contributes to the investment value. For the evaluation of the structuring alternatives we use AHP analysis and perform pair-wise comparison among the alternatives. For the AHP implementation we adopt the commercial tool Expert Choice.
- Perform sensitivity analysis in order to understand the contribution of each criterion in the overall business utility.

Figure 4. The risk mitigation sub-module of the proposed framework



A Case Illustration

To illustrate the proposed methodology we apply it to an ICT investment decision for a growing Water Supply & Sewerage Company, which we refer to as WSSC to protect its identity and its projects. We focus on the risk mitigation factors. The Company's principal business is the supply of water and sewerage services to over 1.5 million people.

The WSSC is challenged on several areas. First, there is an opportunity for the WSSC to offer advanced water management services to its existing customers. This results to enhanced service quality and efficient control of its operating expenses. In addition, its service area is going to significantly increase attracting new customers. In order to achieve all these WSSC management is focusing on the significance of the ICT applications that could transform the company's relationships with customers, suppliers, other partners and environment regulators.

WSSC is interested in proceeding on implementation of an ICT platform in order to improve automation aspects of its operations, decision taken methods, customer services as well as new strategic opportunities in long-term perspective.

A. Define the Content of the Overall ICT Business Activity and its Risk Profile

Analytically, the investment opportunity under examination includes.

- Telemetry-ICT (TICT) infrastructure to enable WSSC to perform more efficiently water network management as well as Asset Management.

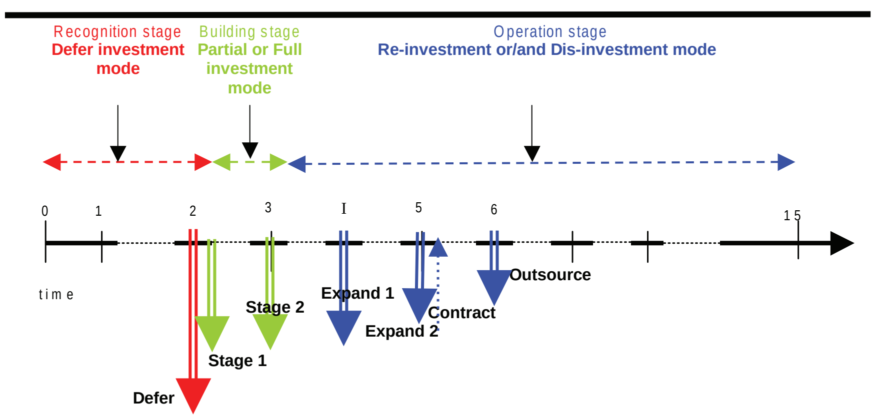
WSSC also examines the possibility of integrating two extra tools into the ICT platform for improving the decision support processes:

- ArcInfo, a Geographical Information System (GIS), that allows users to create, view, access and analyze map (geo-referenced) data.
- StruMap, a Hydraulic Analysis simulation tool, which helps the Water Network Modeling and therefore the Water Management.
- Finally, based on the information collected and manipulated by the afore mentioned platform and its sub-modules a web-based customers support tool will be implemented providing also on line question and answer service to WSSC customers.

Of course, the system integration with the functional internal processes of the company requires re-engineering of the company's internal processes and especially these that are related to the customer support and information.

In our analysis we consider the aforementioned investment opportunity as an overall entity. We focus on Risk Mitigation Factors Perspective.

Figure 5. A configuration of options taken into account in our analysis



Defer: To defer (up to T_d=2years) the base scale TICT platform that contains the Telemetry-ICT system for water management monitoring system including the asset management module (D).

Stage: To build ICTP base scale platform in two stages at $T_{s1}=T_d$ and $T_{s2}=T_d+1$ option to stage 1 (S1) and option to stage 2 (S2). The option to stage S1 concerns the pure telemetry system while the options to stage S2 concerns the asset management sub-module.

Expand: We consider the option to expand (E1) operation at $T_{e1}= T_d+2$ as well as the option to expand (E2) at $T_{e2}=T_d+3$. The first option to expand concerns the GIS and HM sub-modules integration with the Telemetry-ICT base scale system, while the second option the development of a new web-based customers support tool that is based on the technical information retrieved by the already installed and operated sub modules of the ICTP system.

Contract: The company's management may also decide to contract operation for investment stage (E1) instead of expanding them to E2 web-based customer support tool if business conditions become unfavorable.

Outsource: The company's management examines the possibility of outsourcing the base scale operation and maintenance of the T-ICT system after 4 years of operation (i.e. at $T_o=T_d+4$)

The investment goals, requirements, assumptions and risks are summarized as follows.

Risks Presentation

Concerning the specific investment opportunity for the WSSC there are some risks that can influence negatively its performance. In our case, there are mainly company-specific risks, environmental and technological risk since the Company does not experience any customers demand uncertainty or any competition threat. In particular, many of the ICT investments projects either completely fail, or deliver reduced functionality. ICT platform may experience lack of users acceptance, but also lack of companies personnel technical expertise. Also, risk factors may concern unrealistic implementation schedule and environmental complexities such as installation of complex equipment in a large scale that can cause inconvenience to the customers. Risks modeling included in our analysis is given in gray cell in Table 1.

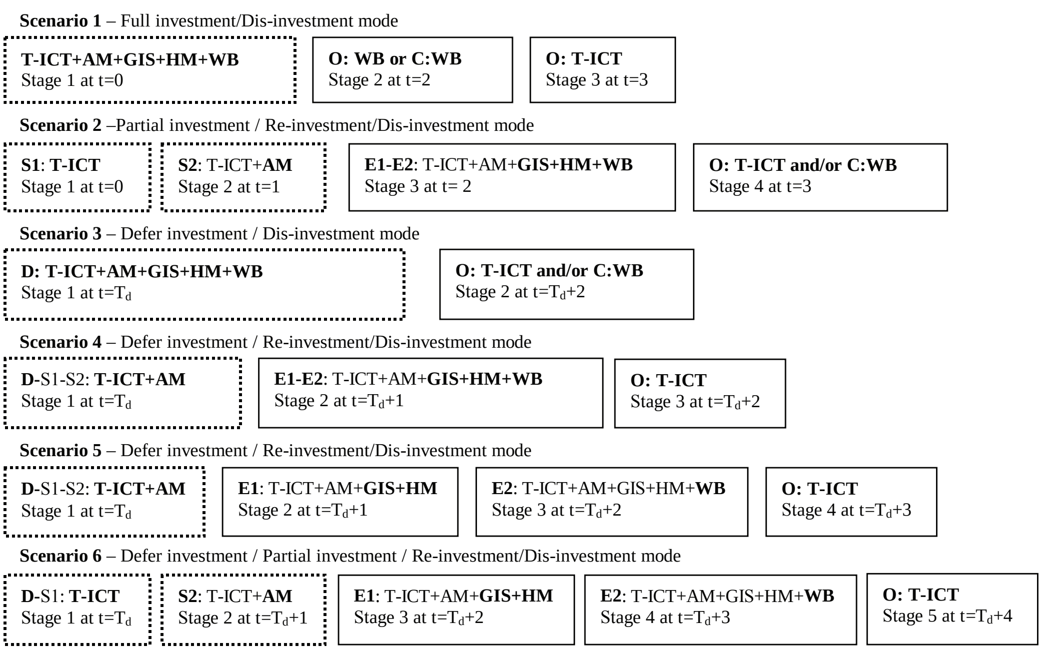
Finally, there is uncertainty about the firm's capability to integrate efficiently the initially planned scale of the ICT infrastructure with the required applications as well as with the content of them.

B. Recognize the options mapped to specific risks and use them to adopt investment modes to be examined.

In this step recognize the options that the investments could embed, based on the identified investment risks. The overall investment deployment strategy in its most complex strategy can include all the options that can control partially or fully the specific investment's risks, Figure 5. In our analysis we consider the options to defer, stage, expand, contract and outsource operations.

The investment's deployment scenarios to be examined are seen in Figure 6. We are looking for the optimum ranking of these scenarios.

Figure 6. Scenarios examined (options combinations – investment content)



The different boxes indicate different investment stages
(--- recognition and building stages, operation stages)

C. *Evaluate investment-structuring alternatives (investment modes) and find a subset of the recognized options that maximally contributes to the investment value*

We have identified the alternative ways to configure the ICT investment using different subsets of the recognized options. For the evaluation of the investment alternatives we do not adopt the quantitative analysis and existing ROs models such as Black-Scholes formula or binomial model. Our intentions was also supported by the interview process with the company's management, which revealed the degree of uncertainty for the various phases of the investment. The company's management expressed the uncertainties level for each investment phase in qualitative way, since it had difficulties in expressing the volatility of the expected value of investments benefits.

Applying the proposed AHP methodology, the pair wise comparison matrices are derived and the relative performance measures are computed for intangible risk factors.

We use the nine-point scale as suggested by Saaty et. el. (1994) however, modified in order to incorporate with our analysis. In particular, we judge our projects portfolio as extreme risk control (E), very strong risk control (VS), strong risk control (S), moderate risk control (M) and equal risk control (E) including intermediate values between the main characterization types. By using the Expert Choice and making judgments according the aforementioned nine-point scale we derive the pair wise comparison matrixes.

Our example is for intuition purpose only and hence we do the scoring alone. Roper-Lowe and Sharp (1990) comment that since it is sometimes difficult to find technical people who can compare options it is necessary for the analyst to learn in detail about each option and do the scoring alone. We play here the analyst role. We select the consistency ratio level, as according to AHP method a consistency ratio must be less than 0.10 to be acceptable (Saaty, 1994).

After making all paired comparisons, for all alternatives, according to the principles of the AHP with respect to all criteria defined in the proposed framework, we compute the total priorities for the alternatives using the known commercial tool Expert Choice. The prioritization results are given in Figure 7.

As it can be seen higher risk mitigation utility value is given by scenario 6 and scenario 5 is following. However, though more option may generate higher control risks in general the financial tangible and intangible factors contribution may change the aforementioned conclusion.

Figure 7. Scenarios prioritization performed with Expert Choice tool

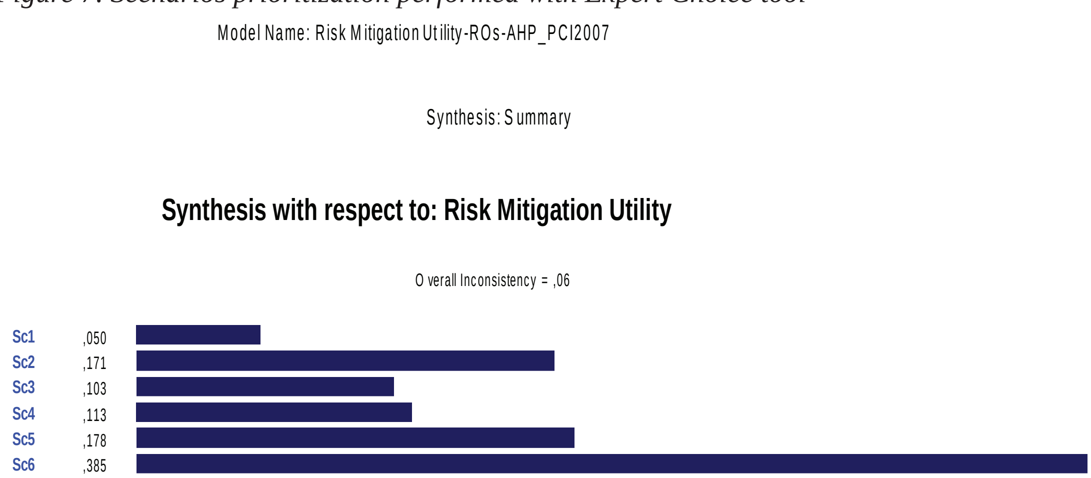
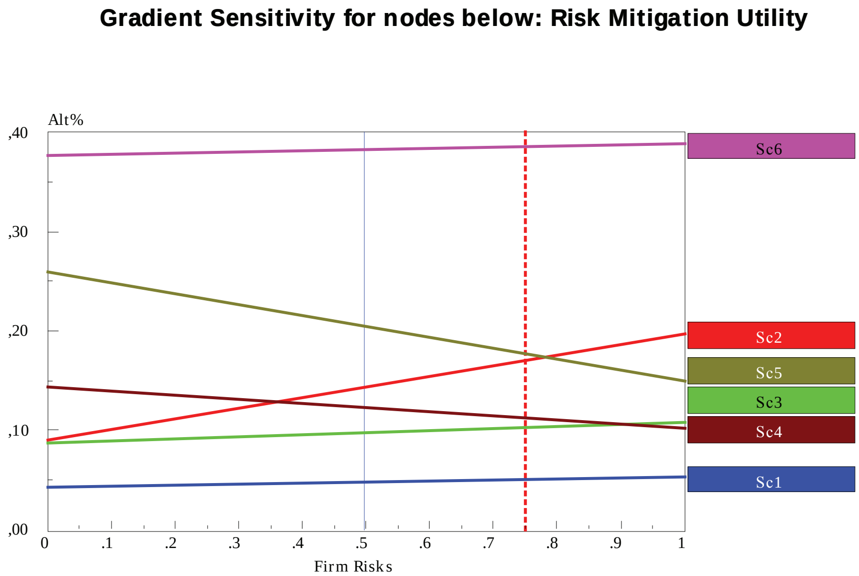


Figure 8. Sensitivity analysis for firm specific risks



D. Sensitivity analysis

Performing sensitivity analysis we can study how sensitive the priorities of the alternatives are to the changes of the input data, i.e. the importance of the criteria.

Figure 8 shows the sensitivity analysis of the results with respect to the importance of firm risk factors.

The figure shows that scenario 6 experiences the highest utility value. Also, for firm risk factors weights from 0 up to 0,75 scenario 5 presents the second priority, while from 0,75 to 1 scenario 2 becomes the second best.

The input data are quite subjective, especially the intangible ones. For this reason, it is important to study the dynamics of the sensitivities carefully. For example, if the importance of one criterion changes significantly, the priorities of between scenarios may change due to space limitations we do not present these results while they are available to the interest reader.

5. DISCUSSION AND FUTURE TRENDS

There is empirical evidence to support the fact that managers who are aware of some options-like ideas do a better job of evaluating and managing risky telecommunication investments. Also, senior finance executives are becoming increasingly aware of the need to view major risky capital investments as options.

ROs have already applied in the literature for risk management and evaluation of telecommunication investments. In practice managers identify risks and formulate strategies that can control these risks in the most efficient way. These strategies may contain option analysis and the target is to map options to risks and find the optimum combination of them in order to achieve the optimal value for the investment opportunity. However, the option analysis experiences criticism concerning the need for the parameters quantification of the option models. The issue becomes even more complicated in telecommunication markets. Particularly, after the markets liberalization competition intensity has been increased dramatically and the players are usually so many that oligopoly models are becoming very complicated to be used in practice. Hence, the quantitative analysis of competition influence in telecom investment opportunities is a very difficult task that requires high-level of mathematical modeling and high number of assumptions while managers and practisians quite often do not "feel comfortable" to adopt it.

Instead, we provide qualitative option thinking. Our analysis requires from the management and business analysts to recognize qualitatively the options, during the lifecycle of an investment opportunity that can at least partially control specific risks. Sometimes deferring an investment may be optimum, while some other times the immediate implementation is the best solution adopting afterwards a step-wise deployment strategy during building and operation period. We propose a decision analysis framework where we compare the various types of adopted options and investment modes for a telecommunication opportunity by adopting AHP analysis. In our analysis, we are not only based on qualitative option thinking but also integrate financial tangible and intangible factors coming by both the nature of the telecommunication investments and the options analysis itself.

Our framework of analysis can be also applied in information technology (IT) service management (e.g. ITIL) and IT governance (e.g. Cobit). Particularly, they deal with issues such as how to provide a good service at the good price. People combine these approaches to the maturity concept. For example, risks P1 to P4 mainly depend on the maturity of the firm on the technologies used in the project, on its current infrastructure which indicates how the project will be integrated. IT service management and IT Governance frameworks are not mutually exclusive and could be combined to provide a powerful IT governance, control and best-practice in IT service management, (Mingay and Bittinger, 2002). Indeed, one can map ITIL process onto the perspective of the standard IT Balanced Scorecard (BSC) as follows.

IT BSC Business Contribution

- Financial Management

IT BSC User Orientation

- Service Level Management
- Availability Management
- Continuity Management
- Incident Management
- Financial Management

IT BSC Operational Excellence

- Problem Management
- Service Level Management
- Change Management
- Service Level Management

IT BSC Future Orientation

- Service Level Management
- Capacity Management
- Change Management
- Financial Management

A similar mapping exists for the CobiT processes. For instance, in the delivery and support domain such as define and manage service levels, manage performance and capacity, ensure continuous service, etc. maps well onto one or more ITIL processes such as service level, configuration, capacity, availability management etc. (Salle, 2004).

Similar to this analysis the aforementioned perspective can be applied to our model considering an AHP hierarchical structure.

6. CONCLUSION

In this work we provide a decision analysis framework for prioritizing telecommunication investment deployment strategies adopting qualitative option thinking. In our analysis, we take into account financial tangible and intangible factors and quantify managerial flexibility in the projects' implementation strategy. So far in the literature, ROs models employ only a quantitative factors analysis for both benefits and costs. However, very often a telecom ICT project also owns a number of qualitative factors that should be taken into account in parallel with the quantitative ones. In addition, the ROs analysis itself in existing models requires input parameters that very often are difficult to be fully quantified. For this reason, we adopt the AHP in order to structure risk profile for telecommunication investments and recognize options that can optimally control them as well as combine tangible and intangible factors into one utility function. It is the first time in the literature where ROs and AHP are combined in a common framework of analysis offering a multi-criteria prioritization model that contains risk, intangible and tangible factors related to telecom business activities.

A limitation of our work is the lack of competition modeling especially focusing on oligopoly cases. In this case, someone could adopt game theory and integrate it with our analysis. In particular, we can consider two players where each of them has to configure its investment in a way that is optimum for him taking also into account the decisions of the competitor.

REFERENCES

Angelou, G. N., & Economides, A. A. (2007A). E-learning investment risk management. *Information Resource Management Journal*, 20(4), 80-104, October-December.

Angelou, G. N., & Economides, A. A. (2007B). A Real Options approach for prioritizing ICT business alternatives: A case study from Broadband Technology business field. *Journal of the Operational Research Society*. Forthcoming doi:10.1057/palgrave.jors.2602477.

Angelou, G. N., & Economides, A. A. (2008). A decision analysis framework for prioritizing a portfolio of ICT infrastructure projects. *IEEE Transactions on Engineering Management, 55* (3), 479-495.

Angelou G., & Economides A. (2005). Flexible ICT investements analysis using real options. *International Journal of Technology, Policy and Management*, *5*(2),146–166.

Benaroch M. (2002). Managing information technology investment risk: a real options perspective. *Journal of Management Information Systems*, *19*(2), 43-84.

Benaroch, M., & Kauffman, R. J. (2000). Justifying electronic banking network expansion using real options analysis. *MIS Quarterly*, *24*(2), 197–225.

Benaroch, M., & Kauffman, R. J. (1999). A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research, 10*(1), 70-86.

Bodin L., Gordon L., & Loeb M. (2005). Evaluating information security investments using the analytic hierarchy process. *Communications of the ACM*, *48*(2), 78-83.

Bräutigam, J., & Esche, C. (2003). Uncertainty as a key value driver of real options. *7th Annual Real Options Conference*.

Expert Choice (1995). *Expert Choice for windows based on Analytic Hierarchy Process,* Version 9, User Manual. Pittsburgh, PA.

Fichman, R., Keil, M., & Tiwans, A. (2005). Beyond Valuation: "Option Thinking" in IT Project Management. *California Management Review, 47*(2), 74-95.

Hallikainen, P., Kivijarvi, H., & Nurmimaki, K., (2002). Evaluating strategic IT investments: An assesment of investment alternatives for a web content management system. *Proceedings* *of the 35th Hawaii International Conference on Systems Sciences.*

Gunasekaran, A., Love, E., Rahimi, F., & Miele, R. (2001). A model for investment justification in information technology projects. *International Journal of Information Management*, *21*, 349-364.

Iatropoulos, A., Economides, A., & Angelou, G. (2004). Broadband investments analysis using real options methodology: A case study for Egnatia Odos S.A. *Communications and Strategies*, *55*(3), 45-76.

Kim J. (1998). Hierachical structure of intranet functions and their relative importance: Using the analytic hierarchy process for virtual organizations. *Decision Support Systems*, *23*, 59-74.

Lai, S., Trueblood, P., & Wong, K. (1999). Software Selection: a case study of the application of the analytic hierarchy process to the selection of multimedia authoring system. *Information & Management*, *36*, 221-232.

Mingay S., & Bittinger S. (2002). Combine CobiT and ITIL for Powerful IT Governance. In *Research Note*, TG-16-1849, Gartner.

Mun, J. (2002). *Real Options Analysis: Tools and techniques for valuing strategic investments and decisions.* Finance: Wiley.

Renkema, T. (1999). The IT value quest: how to capture the business value of IT-based infrastructure. John Wiley & Sons, Ltd.

Roper-Lowe G, & Sharp J. (1990). Analytic hierarchy process and its application to an information technology decision. *Journal of Operational Research Society, 42*(1), 49-59.

Saaty, T. L., & Vargas, L. G. (1994). Decision making in economic, political, social, and technological environments with the analytic hierarchy process. Pittsburgh: RWS Publications.

Sallé, M. (2004). *IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing.* Trusted Systems Laboratory, HP Laboratories Palo Alto.

Schniederjans, M., Hamaker, J., & Schniederjans ,A. (2005). *Information technology investment – decision-making methodology.* Singapore: World Scientific Publishing.

Tam, M., & Tummala, R. (2001). An application of the AHP in vendor selection of a telecommunications system. *Omega*, *29*, 171-182.

Trigeorgis, L. (1996). *Real options: managerial flexibility and strategy in resource allocation*. Cambridge MA: The MIT Press.

KEY TERMS

Analytic Hierarchy Process (AHP): A process that deals in a hierarchical structure a number of quantitative and qualitative criteria in one utility function.

Business Utility: It is the overall utility of the under investigation business activity for the company. It contains both quantitative and qualitative factors.

Deployment Scenario: The way the investment is deployed in terms of time and scale implementation.

Expanded Net Present Value: It is the total value of a project that owns one or more options. It is given by: Expanded (Strategic) NPV = Static (Passive) NPV + Value of Options from Active Management.

Investment Cost: The amount of money spent for the investment, investment expenditure required to exercise the option (cost of converting the investment opportunity into the option's underlying asset, i.e. the operational project).

Real Option: It is the extension of financial options concept to real assets.

Real Option Under Competition Threat: The value of ROs when during waiting period before exercising it there is the threat that a competitor may preempt the owner of the ROs and "steal" or abstract its value.

Risk Management: The process of identifying different types of risks, assessing their relative importance for the project, and implementing strategies for managing risk.

Chapter XXVII
A Business Planning Framework for WiMAX Applications

Rebecca De Coster
Brunel University, UK

ABSTRACT

Mobile networking refers to wireless technologies which provide communications between devices. Applications for mobile networking have a broad scope as they can be applied to many situations in either industrial or commercial sectors. The challenge for firms is to better match market-induced variability to the organizational issues and systems necessary for technological innovation. This chapter develops a business planning framework for mobile networking applications. This framework recognises the fluidity of the situation when trying to anticipate and model emerging wireless applications. The business planning framework outlined in this chapter is a generic model which can be used by companies to assess the business case for applications utilizing mobile networking technologies.

INTRODUCTION TO WIMAX APPLICATIONS IN THE MOBILE NETWORKING SECTOR

WiMax is a wireless standard that was developed to provide a wireless alternative to cable and DSL for broadband access (WiMax-Forum, 2004). The focus here is assessing the potential business case for WiMAX applications in the mobile networking sector. Mobile networking refers to wireless technologies which provide communications between devices, for example, a laptop maybe connected wirelessly to a printer via radio using the Bluetooth standard. The WiMAX standard allows for both point-to-point and point-to-multipoint configurations – the latter being suitable for mobile networking applications.

Point-to-multipoint microwave networks have been previously deployed as proprietary networks (Vaughan-Nichols, 2004), until the development of the Institute of Electrical and Electronics Engineers (IEEE) 802.16 set of standards known as WiMAX (Worldwide Interoperability for Microwave Access). These global standards improve matters in two ways: compatibility of components within a communications system and customer interface standards. The availability of global standards for mobile networking technologies enables high technology firms to focus resources on specific aspects of a WiMAX communication system.

The Business Case for WiMAX Applications

The purpose of the chapter is to examine the forecasting elements for emerging broadband wireless access applications. The chapter examines the elements involved in deploying a new product or service to determine whether or not a business proposal will create value. Value management is an approach to management based on the principle that business decisions are based on the premise that *"they must manage a firm's resources with the ultimate objective of increasing the firm's market value*" (Hawawini and Viallet, 2007, p. 521).

The availability of global standards such as WiMAX is a means for achieving greater adoption. This has been shown with the Wi-Fi (wireless LAN) networking standard where high user adoption rates have been experienced within a relatively short time period. However, significant ROI (return on investment) is not yet clear for WiMAX (Koffman and Roman, 2002) as the network and equipment costs are still high. The breadth of factors which need to be addressed during business planning for broadband wireless access is shown in Table 1.

The ROI components present in Table 1 illustrate the high infrastructure costs involved with investment in both core and access networks with Gunasekaran and Harmantzis (2005) quoting backhaul costs of T3 as $2000 per month and OC3 of $5000 per month. This chapter develops an approach to examining the business case for WiMAX applications and develops a business planning framework. The business planning framework outlined in this chapter is a generic model which can be used by companies to assess the business case for applications utilizing mobile networking technologies.

Table 1. ROI components for broadband wireless access

ROI Components	Key Factors	Key Variables
Revenues		
Applications	Voice/ data/ Internet	Supporting interfaces
Subscription take-up rate	Market Potential	Demographics; capacity per subscriber
Pricing Package	Installation charges and monthly subscription charges	WiMAX billing/tariff structures
Costs		
Spectrum Licence fees	National availability of spectrum	National regulatory situation
Planning costs	Radio planning; business case	Tower site acquisition; frequency allocation and environment (multi-path)
Equipment costs	Backhaul and base station costs	Network topology ; wireless access configuration
Deployment costs	Installation charges	Ease of interconnectivity
Operating costs	Upgrades; maintenance etc..	Equipment interoperability

KEY COMPONENTS OF THE BUSINESS PLANNING FRAMEWORK

The business planning framework for mobile networking technologies which has been developed is shown in Figure 1. The framework comprises two sides which need to be in alignment. The left-hand side identifies the elements to assess the market potential of mobile networking applications. The right-hand side of the framework identifies the elements to assess the business case of mobile networking applications.

This approach differs from other forecasting approaches by utilising an open-systems view of innovation management. This means that the business planning framework supports generic approaches to planning innovation for emerging broadband wireless applications. Each element within the framework is a generic one and does not constrain the other elements in terms of their structure.

This framework recognises the fluidity of the situation when trying to anticipate and model next-generation wireless applications. Previously broadband wireless access was considered an alternate solution to broadband access offered via DSL or cable due to the prominence of the established network operators providing local loop services. This does not take into account the drivers and market development for next-generation wireless applications. The key developments for users were identified as working assumptions for fourth-generation mobile project (Bria et al, 2001) including: telepresence (in place of meetings); information anywhere, anytime; intermachine communication; security and one-stop shopping. Many consumer electronic applications in the mobile networking sector arise due to user demands for the convenience of wireless connectivity.

The business planning framework outlined in Figure 1 is a generic model which identifies key components by which business proposals can be examined and the potential value identified.

Key Components of Market Potential

The business planning framework recognises that firms must assess market potential to maximise return on investment (ROI). This requires analysis of collected information concerning market structures and market offerings in a systematic manner. An approach to assessing the market potential for broadband access services is to focus on the different demographics and the WiMAX billing/tariff structures as illustrated by a couple of

Figure 1. The business planning framework for mobile networking technologies

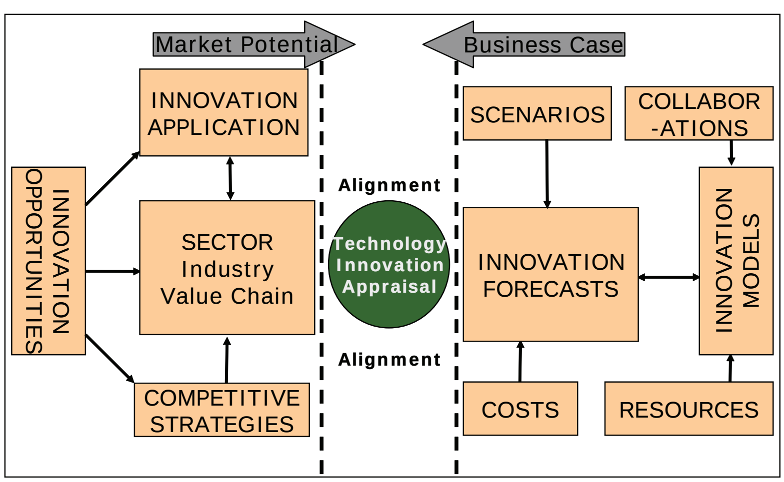


Table 2. Modelling the market potential of broadband wireless access

Model Assumptions	Residential Example: Low Usage	Residential Example: High Usage
Customers	Residential	Small office
Customer capacity requirement	2MB/s to 5MB/s	5MB/s to 10MB/s
Geographic coverage	Depends on allocated frequency and radio planning	Depends on allocated frequency and radio planning
Customer density	Number of homes in area	Number of offices in area
Initial Take-up estimates	20%	30%
Target Population	30%	40%

example scenarios in Table 2. The models shown are extensions on existing models for broadband wireless access (Petkovic and De Coster, 2000) including capacities forecasted for fourth and fifth generation mobile systems (Ohmori et al, 2000 and Ghosh et al, 2005) rather than capacities currently reported (Ballon, 2007).

Revenues are essentially based on subscription take-up rates multiplied by the pricing package. The nature of the customers will be the key variable as residential users will have very different requirements and pricing levels to corporate users. Hence, demographics are a key factor in determining the business viability of any telecoms network (Ballon, 2007).

Pricing components of a broadband wireless access service usually comprise two elements (Petkovic and De Coster, 2000): installation charges and monthly subscription charges. The installation charges need to cover the logistical and personnel costs and may include equipment costs. The monthly subscription charges are likely to consist of a mix of flat-rate or packaged WiMAX billing/tariff structures depending on the type of customer.

Innovation Opportunities

The successful commercialisation of products encompassing new technologies requires a high level of market sensing (Anderson and Narus, 2004, p41). This refers to the ability of firms to anticipate the desires of customers and trends in markets. The fast moving high technology sector of mobile networking is one where establishing market sensing is not easy to do. Many of the protocols and standards are only recently established and are yet to be deployed in the marketplace, for example, mobile WiMAX could support data transfer at vehicular speeds (Vaughan-Nichols, 2004).

Wi-Fi has been shown to be a cost-effective broadband wireless access solution for rural and remote areas (Zhang and Wolff, 2004). The success of Wi-Fi was explained by three factors (Galperin, 2005):

1. Provides wireless access as well as for backhaul traffic;
2 Widespread industry support for the standard and
3 The lack of regulatory overhead.

The growth in diversity of mobile devices is yet to occur and includes Smartphones, PDAs, Portable Media Centers, retail point-of-sale systems, Global Positioning System-based devices and industrial robots. This increasing number of user devices are challenging to support as they each have different interface requirements. The end users are seeking to gain improved usability through wireless connectivity. WiMAX has been added to the list of approved 3G radio access technologies in the International Mobile Telecommunications 2000 (IMT-2000) standard as defined by the International Telecommunication Union (ITU).

It is unclear how attractive the services offered by third-generation mobile (3G) networks will be compared to Wireless-LANs such as Wi-Fi (Lehr and McKnight, 2003). From the end-user's perspective value for the end-user is the means to express in monetary terms the functionality or performance of a market offering in a given customer application (Anderson and Narus, 2004). For example, a home cordless base station will connect a mobile phone to the public telephone network in order to reduce cost of calls and remove the need for dedicated home wireless terminals as cordless phones.

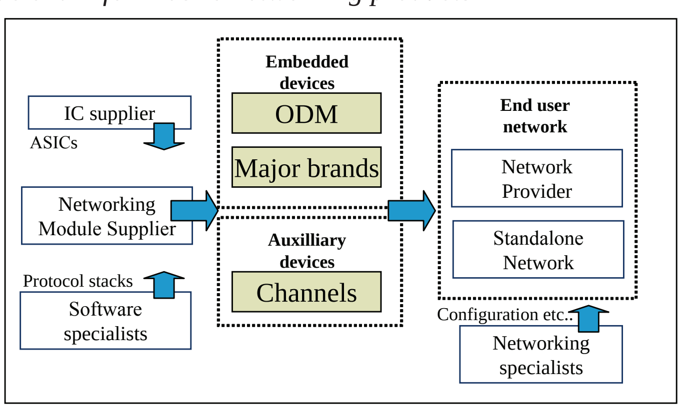
Consumer applications utilising wireless connectivity are increasing including multimedia communication to PC peripheral devices including Digital Still Cameras; Camcorders, MP3 players and mobile phones. Applications for mobile networking have a broad scope as they can be applied to many situations in either industrial or commercial sectors. The take-up will depend on the perceived benefits of future applications and whether the differences in quality of service will overcome higher user charges. For example, the quality of service for a voice-over-IP which can be offered over Wi-Fi connecting to an IP network versus the high quality of a voice service over 3G networks (Laroia et al, 2004 and WiMax-Forum, 2007).

The Mobile Networking Industry Value Chain

The approach taken in the business planning framework is to establish the focus and scope of business activities (Afuah and Bahram, 1995), with respect to the positioning of the firm in the industry value chain. The industry value chain for the mobile networking sector is shown in Figure 2. It comprises many high technology companies which have specialist expertise and manufacturing capabilities.

The industry value chain for the mobile networking sector comprises three elements from upstream to downstream as follows: vendors, technology providers and the end user network as shown in Table 3. Firms have developed their mobile networking expertise in-house or externally, for example, acquiring specialist technology companies with audio recognition software that improves the audio performance of any voice-based product or system. Many firms use specialists to provide network monitoring equipment aimed at data acquisition for the purposes of remote monitoring including the monitoring

Figure 2. Industry value chain for mobile networking products



of equipment operation at unmanned sites or as automated disaster prevention systems.

The shortening of product lifecycles increases pressures on the management of product innovation activities to reduce time-to-market. Specialising within the industry value chain firms is a means for achieving this for high technology firms operating in highly uncertain markets such as mobile networking where the applications are still emerging (Bria et al, 2001).

The scope of activities that a firm provides within the industry value chain is part of the strategic focus of a firm (Afuah and Bahram, 1995). The companies specialize within the industry value chain for mobile networking products which comprises a number of high technology firms each developing technologies and products for deployment in international markets. The advantages of this approach (of having a devolved value chain), is that firms are utilising reduced transaction costs and economies of scale (WiMax forum, 2004). This is particularly relevant for mobile networking applications where solutions encompass a variety of technological approaches in terms of wireless encoding, data communication and networking protocols.

The other benefits of having the value chain being provided by different organisations is that the business is more flexible as, unlike a large corporation, the corporate structure is more easily changed (Floricel and Miller, 2003). The organisation is also more responsive to changes in market conditions as each of the various firms in the network is monitoring the market (Stuart, 2000).

Table 3. Summary of the elements of the mobile networking industry value chain

Upstream Elements	
IC (Integrated Circuit) Suppliers (including ASICs – application specific ICs)	ASICs and other specialised ICs are a key part of mobile networking products and the companies with the design capability for this may not necessarily manufacture the ICs themselves
Networking Module Suppliers	Firms that provide the hardware components including the circuit board, antennas, transceivers and base process units
Software Specialists	Firms that provide the software necessary to handle the end user application (or interface to an existing application) plus the communications and networking functionality required
Midstream Elements	
Original Device Manufacturers (ODMs)	Firms who supply electronics equipment that they require to become wireless enabled
Major Brands	The major electronics companies – typically global consumer electronics companies with an established market brand
Channels (to market)	These are the distribution channels providing mobile user devices to consumers (e.g. high street retailers)
Downstream Elements	
Network Providers	These are typically large service providers (including mobile phone operators)
Standalone Networks	Firms providing industrial networks (e.g. in oil refineries)
Networking Specialists	These are typically consultants with expertise in the configuration, installation and operational aspects of communication networks (including technical and other performance issues)

Innovation Application

Market driven management starts with the definition of the target customers and their applications. This concurs with the view taken here since a firm's focus on specified target customers will define the parameters of the product/ service offering. Consumer products are characterized by large volume and very tight cost pressures on the producer. To achieve a comparative advantage requires technology firms to develop their technology understanding prior to volume market launch. They then utilize their market knowledge to launch desirable product configurations (Orihata and Watanabe, 2000).

The chosen innovation applications are often dependent on a firm's legacy, its core competences, and its reasons for moving into the mobile networking sector and recent product development activities. This focus on specific applications is necessary to establish the functionality provided by the mobile networking products and optimise the mobile networking solution to the end application.

WiMAX uptake may benefit from the increasing number of mobile user devices which are available both for industrial purposes (such as handheld terminals) and for consumer electronics. Healthcare is an area where we are likely to see a proliferation of "always-on", battery-powered devices both in the home and in hospitals. Home networking is an area which has been recognised but has not yet seen great diffusion. This will entail linking devices together and exchanging information so that a high level of automation is achieved, hence, driving the potential demand for WiMAX.

Competitive Strategies

The strategic orientation of a firm in highly uncertain markets requires three areas of focus according to Gatignon and Xuereb (1997): customer; competitors and technological. The approach taken in the business planning framework is to evaluate a firm's competitive performance in the mobile networking sector on the three aspects of: resources; market position and financial performance. These factors were identified by Hunt (1999) in the work developing the RA (resource-advantage) theory of competition. The reasons for adopting this theoretical framework is that it combines the literature on the resource-based view (RBV) of the firm as well as the literature on strategic positioning or market based view (MBV).

As technology advances it becomes increasingly difficult for firms to have sufficient resources of sufficient breadth and depth in the required technological areas. This necessitates the need to create a business model involving collaboration as technology firms are driven by resource limitations. The importance of business models is that they can assist firms in their quest for developing competitive advantage on an ongoing basis.

As technology becomes more common, competition may shift away from technology to other areas (Friar and Horwich cited in Lemos and Porto, 1998). WiMAX as a mobile networking application enables equipment providers to have a much closer relationship with the end users of their broadband wireless access service. There is the opportunity for technology firms to gain insights into user aspects which enable them to improve their future product offering in terms of interactive design and other user-centred design aspects. This knowledge can provide a firm with competitive intelligence that should enable the firm to maintain a competitive advantage (Lemos and Porto, 1998).

Key Components of the Business Case

A business case assessment should identify the proposal's potential contribution to the overall performance of the firm (Loch, 2000). Business planning involves understanding the drivers of value both for the firm creating the product/ service and the customer purchasing it (Ryan, 2004, pp. 81). From the firm's perspective the business case should examine whether the value of the future stream of net cash exceeds the initial cash outlay required for a business proposal (Hawawini and Viallet, 2007, p.5).

Business models should not be restricted to pricing and revenue models but should also address the following according to Osterwalder and Pigneur (2002): the value proposition of the proposed product innovation; customer management; infrastructure management and financial aspects. Infrastructure management comprises the following according to Osterwalder and Pigneur (2002): capabilities and resources; value configuration and partnerships. This is reflected in Figure 1 with the "innovation models" allowing for internal resources as well as collaborations.

Innovation Models

Innovation models identifies the approach a firm takes to organising itself for undertaking the development work necessary to develop the new products/ services. Essentially it is a "business model" for the firm's activities relating to innovation which describes the key development activities. Technology core competencies for mobile networking firms are multi-disciplinary. Existing technology firms are unlikely to have developed expertise in all the required areas. To develop mobile networking products and services they need to identify the expertise that should be developed to support the development of the required products (McEvily et al, 2004).

It is necessary to establish the costs of the technological innovation and product development activities which firms have undertaken to develop products in the mobile networking sector. The trend is towards product development using business partnerships with other vendors who have hardware expertise and product portfolios that will enable a firm to realise their technology development strategy. This is increasingly necessary as mobile networking firms encounter a wider range of consumer electronic scenarios including streaming CD-quality audio; digital image transfer and laser printing where bandwidth and power are critical factors.

Collaborations

Business partnerships are a fundamental change and affect the way in which businesses are organised and managed. Companies are able to retain within the organisation only the activities or processes that best meet their competencies and provide optimum value-add. The reduced resources are focused on a given activity (or set of activities) that are revised and refined till a high degree of sophistication and speed is achieved.

The business case framework shown in Figure 1 recognises that it is common practice to utilise business partnerships and collaborations during development work. This is particularly true for end user mobile devices where the shortening of product lifecycles increases pressures for innovation activities to reduce time-to-market. Fast development processes enable firms to exploit the emerging innovation opportunities which arise from the interaction among components once new technologies are in place. These new innovation opportunities were not apparent prior to launch of the new technologies and result from the interactions. This is apparent in mobile networking as more and more user applications are being realised after systems have been deployed in the field.

Innovation strategies that involve other firms requires an organisational commitment to the pursuit of a chosen product development with another organisation. Despite this alliances have been recognized as providing efficiency and greater creativity for new product development. Both of these are key to addressing the issue of sustainability of competitive advantage. Faster renewal of resources and products is achieved which is necessary to counter changes of markets, competitors and technological advancements.

Resources

Ryan (2004, p.272) identified that the financial planning for a business case should comprise three parts:

1. The revenue model which describes how the new business activity will win revenue
2. The expenditure model which describes the necessary expenditures required to achieve the goals specified by the revenue model
3. The financing model which describes how the firm will raise its finance

Forecasts are then created based on the projections made for these three parts. The forecasts should consider the value proposition over the entirety of the value life cycle which according to Osterwalder and Pigneur (2002), comprises: value creation; purchase; consumption, renewal or transfer. The scenario for the provision of broadband wireless access is rather different as user costs are predominantly service access charges (e.g. subscriptions); rather than the equipment purchased. The business case is altered so that the revenue is generated over the lifetime usage of the equipment rather than the initial purchase price. This approach focuses on the service provider providing end users connectivity to the Internet via broadband wireless rather focusing on the mobile user devices.

The two items of interest here are the revenue model and the expenditure model which, according to Ryan (2004, p.272); are dependent on the internal constraints of a firm and external opportunities that the business confronts as shown in Table 4.

Innovation Forecasts

Forecasting essentially requires an understanding of the broad area beforehand (Cuhls, 2003). This more general establishment of the future context is known as foresight and often starts with the identification of the one or the different options for the future. By nature foresight is more qualitative that quantitative and usually precedes detailed forecasting which is more quantitative in nature.

The qualitative approach to financial forecasting has limits (Kesh and Raja, 2005), but can be adopted for specific applications. The approach requires a high level of clarity of terms and establishing a relationship amongst the elements. The relationship could be a hierarchical arrangement which defines the necessary logic behind the elements in a model for financial forecasting (Lemos and Porto, 1998). The experiences of the electronics sector in providing mobile handsets in large volumes to short market windows is unlikely to be the pattern for WiMAX adoption.

A switch to mobile networking applications will not only involve potential forecasting difficulties in terms of market research amongst WiMAX end users (McBurney et al, 2002) but also organisational forecasting problems (Sand-

Table 4. Internal and external factors affecting the business case

The revenue model:	
External revenue drivers:	Markets, competition etc..
Internal revenue drivers:	Capabilities and USPs etc..
The expenditure model:	
External cost drivers:	Supply chain characteristics, labour conditions etc..
Internal cost drivers:	Organisational structures, technologies in place etc..

ers, 1995) amongst the network operators who are not used to this business approach. Resistance to adopting new financial forecasts is likely to occur until such a time that field trials enable forecasts to be evaluated (Sanders, 1995). The difficulties of forecasting broadband wireless access service provision are due to the lack of established applications and that timescales for the forecasted positive net cash flow are lengthy. One approach to overcoming organisational forecasting problems is to develop co-operative strategies (Sanders, 1995).

Scenarios

Financial planning requires quantifying the expected cash flows for the innovation opportunities. As WiMAX mobile networking applications are not yet deployed it is not possible to measure cash inflow from operations. The implication for examining the business case for WiMAX applications is that one of a number of alternative valuation methods will need to be proposed. Scenarios are a method of approaching this by recognising a potential future state which can then be evaluated financially by means of forecasting. Forecasting market demand for new telecommunications services cannot make use of historical market data which makes scenario analysis methods an attractive approach to modelling the dynamics of a marketplace (McBurney et al, 2002).

WiMAX is likely to benefit from both increased capacity demand from the industrial sector as well as the consumer sector which has a large number of consumer devices which are Wi-Fi enabled. Industrial applications range from temperature monitoring, security, industrial monitoring and building automation markets through to medical, automotive, retail (particularly electronic point of sale) and factory automation. It is difficult to assess potential user interest in the case of technology that is novel and largely unknown to the market. While the provision of Internet services is well established, the value to customers of telecoms services being delivered by WiMax is unclear at present. The end user will need to be reassured that emergency delivery of calls is guaranteed and that adequate technical support is available.

Costs

Quantifying the costs of deploying WiMAX mobile networking applications requires understanding the different costs involved. The categorisation of costs was identified earlier (in Table 1) as:

- Spectrum Licence fees
- Planning costs
- Equipment costs
- Deployment costs
- Operating costs – this refers to ongoing network support costs and will not be discussed further here.

Spectrum licence fees are potentially prohibitive costs in terms of the business case. Regulatory constraints are a factor for WiMAX as the availability of radio spectrum is a regulatory constraint. Much of the spectrum needed to deploy WiMAX has already been distributed by governments or dedicated for other purposes by network operators (Vaughan-Nichols, 2004).

Planning costs are not the issue in themselves as it usually involves a radio planning group examining the geographic location and determining performance parameters. The issue is more the implications of the findings of the planners as these will impact the amount of equipment required to be deployed. The performance of the radio transmissions are affected by the radio frequency which in turn affects radio planning, for example, in determining cell capacity (Fong et al, 2004).

Equipment costs vary depending on how closely tied a service provider is to a vendor. Proprietary wireless solutions incur higher equipment costs as equipment vendors have to recover substantial development costs. In contrast, global standards enable development costs to be spread across the wider community of high technology companies. Further, the overall adoption rate is slower than established standards such as Wi-Fi (wireless LAN), particularly in terms of equipment such as access points and end user devices.

The methodology for forecasting costs for next-generation wireless applications is based on a mixture of three considerations:

1. Historic cost
2. Modelling of equipment
3. Equipment cost decline trends

Historic cost forms the basis on which telecoms networks are costed. The experience of deploying and running a network provides the baseline information for service providers on which they formulate future cost assumptions. In contrast; establishing the amount of equipment to be deployed is based on the anticipated capacity requirements. This provisioning of equipment is in turn is based on forecasts for the take-up of broadband wireless access services.

The business case for broadband wireless access in areas where there is no established access network (e.g. developing countries or rural areas), is likely to differ to the business case for traditional access networks that incurred high costs of core switching equipment and access network deployment. However, backhaul costs are still expensive (Gunasekaran and Harmantzis, 2005). The future IP (Internet protocol), based core network will use a variety of access technologies to connect users (Becher et al, 2001). The costs of communications equipment for the core network are dropping and the components for end user devices are evolving into becoming multifunctional. The IP products in the core network are likely to be provided by other specialist organisations who would supply them to the service provider. For example, application developers, device manufacturers, system integrators and application service providers.

Deployment economics are a key concern for network operators (Koffman and Roman, 2002) as visiting customer homes to install and configure equipment is a resource intensive business. Recent laptops are now fitted with Wi-Fi access cards and users are able to setup and configure a wireless LAN in their home themselves. This reduces the deployment costs and also speeds up roll-out of broadband wireless access on a large-scale. At present most user devices do not have a WiMAX radio interface card fitted. The future trend is likely that radio interfaces in mobile Internet and consumer electronic devices will support several radio interfaces and possibly have the potential to download upgrades (Becher et al, 2001 and Ohmori et al, 2000).

FUTURE TRENDS

Business planning for mobile networking applications needs to consider the perspectives of the equipment providers; the service providers and the end-user. The equipment providers typically incur high development costs and the associated operating expenditure model during production. In contrast, the service providers are concerned with network configuration, deployment economics and operating costs.

There are a number of areas where mobile networking applications will offer new opportunities, including e-Home and in-car information and entertainment systems. Social networking is (Alam and Prasad, 2007): *"A platform where people from all walks of life come together to express themselves by means of sharing videos, pictures, contents, etc.. and it provides the ability to collaborate using peer-to-peer applications and services"*. The trend towards social networking is potentially a global one which could alter the dynamics of capacity utilisation amongst residential users of broadband wireless access.

Financial planning requires quantifying the expected cash flows for these market opportunities. The business case for WiMAX applications needs to be flexible enough to evaluate mobile user services for which the sales growth rate is undetermined. As WiMAX mobile networking applications are not yet deployed it is not possible to measure cash inflow from current operations.

CONCLUSIONS

The chapter started with a description of the potential of a WiMAX communication system to provide mobile networking services. This utilises the point-to-multipoint configuration of the WiMAX standard to provide a broadband wireless access solution. In the context of value management with its emphasis on creating and delivering value, a business planning framework has been introduced. This comprises two sides which need to be in alignment. With respect to examining the market potential the market requirements are examined along with competitive positioning. With respect to the business case a discussion of innovation models was made that recognises the various dimensions of innovation management and resourcing. Concerning the forecasting of future demand for emerging broadband wireless applications the approach proposed is scenario based that recognises the globalisation of services and applications such as Face Book.

The presented business planning framework for mobile networking technologies provides a means of matching market-induced variability to the organisational issues and systems necessary for technological innovation. The framework draws on principles grounded in financial management, marketing management strategy and innovation management. By utilising an open-systems view of innovation management in the business planning framework it supports generic approaches to planning innovation for emerging broadband wireless applications. Further it recognises the challenges of providing quantitative forecasts when forecasting demand for emerging services for which historical data is not available.

REFERENCES

Afuah, A. N., & Bahram, N. (1995). The hypercube of innovation. *Research Policy*, *24*(1), 51-76.

Alam, M., & Prasad, N. (2008). Convergence transforms digital home: techno-economic impact. *Wireless Personal Communications*, *44*(1), 75-93.

Anderson, J. C., & Narus, J. A. (2004). *Business market management – understanding, creating and delivering value*. Upper Saddle River, NJ: Pearson Education, Inc.

Ballon, P. (2007). Changing business models for Europe's mobile telecommunications industry: The impact of alternative wireless technologies. *Telematics and Informatics*, *24*(3), 192-205.

Becher, R., Dillinger, M., Haardt, M., & Mohr, W. (2001). Broadband wireless access and future communication networks. *Proceedings of the IEEE*, *89*(1), 58-75.

Bria, A., Gessler, F., Queseth, O., Stridh, R., Unbehaun, M., Wu, J., & Zander, J. (2001). 4th-generation wireless infrastructures: scenarios and research challenges. *IEEE Personal Communications*, *8*(6), 25-31.

Cuhls, K. (2003). From forecasting to foresight processes - new participative foresight activities in Germany. *Journal of Forecasting*, *22*(2-3), 93-111.

Floricel, S., & Miller, R. (2003). An exploratory comparison of the management of innovation in the New and Old economies. *R&D Management*, *33*(5), 501-525.

Fong, B., Ansari, N., Fong, A. C. M., Hong, G. Y., & Rapaji, P. B. (2004). On the scalability of fixed broadband wireless access network deployment. *IEEE Communications Magazine, 42*(9), S12- S18.

Galperin, H. (2005). Wireless networks and rural development: opportunities for latin america. *MIT Press Journal of Information Technologies and International Development, 2*(3), 47-56.

Gatignon, H., & Xuereb, J. M. (1997). Strategic orientation of the firm and new product performance. *Journal of Marketing Research*, *34*(1), 77-90.

Ghosh, A., Wolter, D. R., Andrews, J. G., & Chen, R. (2005). Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential. *IEEE Communications Magazine, 43*(2), 129-136.

Gunasekaran, V., & Harmantzis, F. C. (2005). Affordable infrastructure for deploying WiMAX systems: mesh v. non mesh. *IEEE 61st Vehicular Technology Conference*, (5), 2979- 2983.

Hawawini, G., & Viallet, C. (2007). *Finance for executives – managing for value creation.* Mason, Ohio: Thomson South-Western.

Hunt, S. D. (1999). *A general theory of competition: resources, competences, productivity, economic growth*. London: Sage Publications.

Kesh, S., & Raja, M. K. (2005). Development of a qualitative reasoning model for financial forecasting. *Information Management & Computer Security, 13*(2), 167-179.

Koffman, I. & Roman, V. (2002). Broadband Wireless Access Solutions Based on OFDM Access in IEEE 802.16. *IEEE Communications Magazine*, *40*(4), pp. 96-103.

Laroia, R., Uppala, S., & Li, J. (2004). Designing a mobile broadband wireless access network. *IEEE Signal Processing Magazine, 21*(5), 20- 28.

Lehr, W., & McKnight, L. W. (2003). Wireless internet access: 3G vs. WiFi? *Telecommunications Policy, 27*(5-6), 351-370.

Lemos, A. D., & Porto, A. C. (1998). Technological forecasting techniques and competitive intelligence: tools for improving the innovation process. *Journal of Industrial Management & Data Systems, 98*(7), 330 – 337.

Loch, C. (2000). Tailoring product development to strategy: case of a European technology manufacturer. *European Management Journal, 18*(3), 246-258.

McBurney, P., Parsons, S., & Green, J. (2002). Forecasting market demand for new telecommunications services: an introduction. *Journal of Telematics and Informatics, 19*(3), 225-249.

McEvily, S. K., Eisenhardt, K. M., & Prescott, J. E. (2004). The global acquisition, leverage, and protection of technological competencies. *Strategic Management Journal*, *25*. 713–722.

Ohmori, S., Yamao, Y., & Nakajima, N. (2000). The future generations of mobile communications based on broadband access technologies. *IEEE Communications Magazine, 38*(12), 134- 142.

Orihata, M., & Watanabe, C. (2000). Evolutional dynamics of product innovation: the case of consumer electronics. *Technovation, 20*(8), 437-449.

Osterwalder, A., & Pigneur, Y. (2002). Business models and their elements. Paper presented at *the International Workshop on Business Models*, Lausanne, Switzerland.

Petkovic Y., & De Coster, R. (2000). *Broadband Access: New Business Models.* London: Ovum Ltd.

Ryan, B. (2004). *Finance and accounting for business.* London: Thomson Learning.

Sanders, N. R. (1995). Managing the forecasting function. *Journal of Industrial Management & Data Systems, 95*(4), 12 – 18.

Stuart, T. E. (2000). Interorganizational alliances and the performance of firms: a study of growth and innovation rates in a high-technology industry. *Strategic Management Journal, 21*(8), 791-811.

Vaughan-Nichols, S. J. (2004). Achieving wireless broadband with WiMax. *IEEE Computer, 37*(6), 10- 13.

WiMax-Forum. (2004). *Regulatory position and goals of the WiMAX forum*. Retrieved May 14, 2008, from http://www.wimaxforum.org/news/downloads/WiMAX_Forum_Regulatory_Whitepaper_v06162004.pdf.

WiMax-Forum. (2007). *IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access*. Retrieved May 14, 2008, from http://www.wimaxforum.org/news/downloads/WiMAX-Whitepaper.pdf.

Zhang, M., & Wolff, R. S. (2004). Crossing the digital divide: cost-effective broadband wireless access for rural and remote areas. *IEEE Communications Magazine, 42*(2), 99- 105.

KEY TERMS

Business Case: Development of a business proposal including the evaluation of resources and finances including the assessment of the likely earning potential of new products or applications.

Business Planning: The process of identifying likely cash flows and examining the sensitivity of these likely cash flows from which targets are established for managing the business development including resourcing.

Forecasting: Preparing predicted customer and market take-up of specific applications which form the basis on which investment criteria are examined.

Mobile Networking: Wireless technologies which provide communications between devices, for example, a laptop maybe connected wirelessly to a printer via radio using the Bluetooth standard.

Scenario Planning: An approach to forecasting which examines likely market situations (or scenarios) based on qualitative research into users and markets which can then be extrapolated into quantitative forecasts.

Chapter XXVIII
Adoption of VoIP Applications in Public and Private Organizations

Nicoletta Corrocher
Bocconi University, Italy and University of Tromso, Norway

Roberto Fontana
University of Pavia, Italy and Bocconi University, Italy

Claudia Parlanti
Bocconi University, Italy

ABSTRACT

We provide evidence on the determinants of diffusion of Voice over IP (VoIP) applications in a sample of public and private organizations in Italy. We first review the recent developments in the technology in order to identify the current trends and the costs and benefits of VoIP adoption. Second, we discuss the recent policy efforts at the European level toward the implementation of the technology. Third, we present an empirical investigation. Our results suggest that organizations become more likely to adopt as time goes by, and that the decision to adopt is mostly affected by size and availability of financial resources. Organizations can benefit from IP communications systems, because they offer cost savings and enhanced applications facilitating network management and on-line transactions. However, technical shortcomings, established habits and practices, and legacy network investments can inhibit adoption. This explains why firms are more likely to adopt as time passes and why small organizations are more inclined to adopt than larger ones.

INTRODUCTION

The growth of the Internet Protocol (IP) has raised the issue of a possible migration from traditional circuit-switched networks designed for basic voice communication towards shared packet transport handling a mix of applications. A substantial number of different services and solutions are currently available for businesses, such as advanced IP telephony calling and management, Web, audio and videoconferencing, instant messaging, calendar and other PIM functions, email, fax, and voicemail. IP Communication lowers costs of network management, since it eliminates the need for different infrastructures. It simplifies maintenance and allows quickly reactions to the changing needs of businesses. However, cost reductions are just one part of the story. Indeed, IP-based networks enhance business communications by providing a flexible foundation upon which all types of new applications and services can be deployed, quickly and easily. Solutions like IP telephony, unified messaging, and IP contact centre have been designed to support increasingly mobile workforce. For instance, unified messaging solutions deliver any message into a single inbox, allowing employees to access and manage their communications with any device. These applications improve communications by facilitating increased mobility, delivering advanced functionality, and streamlining administrative tasks. As a result, employees are able to communicate more effectively and can focus on activities that create new revenue streams or generate cost savings. However, although the technology has been constantly improved, there remain some technical problems for the provision of IP Communications, which are related to the characteristics of the IP network, such as voice quality and network reliability, legacy investments, and regulation.

This chapter has three aims. First it will provide a review of the recent developments in the technology in order to identify the current trends and their impact on the costs and benefits of Voice over IP (VoIP) adoption. Second, it will briefly discuss the recent policy efforts, especially at the European Union level, toward the implementation of the general framework approved in 2003. Third, on the basis of these premises, it will present an empirical investigation of the determinants of the adoption of VoIP Communication by public (i.e. Universities, Hospital, Public Administration offices) and private organisations (i.e. Banks and other types of enterprises), in order to identify drivers and obstacles of the diffusion of this technology. In particular, we intend to understand what firm-specific and market-specific variables drive individual adoption. The empirical contribution of the chapter relies upon an original dataset including 123 Italian organisations that have adopted VoIP Communication solutions since from 2001 (when VoIP Communication was first introduced in Italy) until 2007.

BACKGROUND: THE CHARACTERISTICS OF VOIP TECHNOLOGY

Voice over IP services refer to the transmission of voice and data over the IP network. This application may be also implemented over any form of the IP network, including Local Area Networks and corporate Intranets. VoIP converts the voice from analogue signals to a series of digits, bundles the data into packets and transmits these packets over the network. Early Internet phones evolved either as by-products from the videoconference industry or from companies like VocalTec, the first firm to use Internet telephony in a computer application in 1995. This service was initially available only to users with a computer connected to the Internet. However, following the continuous technological evolution, the cost saving benefits of Internet telephony have become available to any user with a telephone connected to the public switched telephone network (PSTN) (Babbage et al., 1997; Ono and Aoki, 1998).

There are three classes of Internet telephony: PC-to-PC; phone-to-PC (or PC-to-phone); phone-to-phone (Clark, 1997). In the PC-to-PC configuration, both computers are linked directly to the Internet and voice packets travel entirely over a packet-switched network. The phone-to-PC and PC-to-phone models of Internet telephony require the installation of a gateway which links the Internet to the PSTN and sets up a connection with a remote gateway at the other end of the call, in order to ensure compatibility between the two networks. Outgoing calls from a PC to a phone are simpler than calls travelling in the opposite direction, since it is relatively easy for an Internet phone application to find a phone number and the IP address of the gateway, while it is impractical to require the PSTN users to dial first the telephone number of a gateway and then the numeric IP address of the called party (Babbage et al., 1997; Kuprinski, 1997). Phone-to-PC users receive regular phone calls during an Internet session. A window notifies an on-line user of an incoming call and they take the call over the PC while remaining connected to the Internet. The third configuration of Internet telephony is phone-to-phone. In this type of model, the caller's phone is connected to a gateway which transforms voice into IP packets and sends them over the Internet to another gateway. This gateway converts the digital signals into analogue signals and routes the call through the local PSTN, which in turn transmits the call to the telephone of the called party.

VoIP services are also being used by corporate users as an alternative means to enable voice communications (OECD, 2006). There are mainly three types of corporate VoIP services: *gateway type*, which adds gateways to a traditional private branch exchange (a telephony system to switch calls between corporate users on local lines while allowing all users to share a certain number of external telephone lines); *IP-PBX type*, which permits all controls for switching calls implemented by PBX to be managed over the IP; *IP Centrex type*, which implements outsourcing installation, operation and management of IP networks to telecommunications operators. This notion is very broad and includes both consumer and business application. The former implies point-to-point and point-multipoint communications; the latter only multipoint-to-multipoint ones.

The development of broadband and wireless connections has led to a wide diffusion of the Internet and, consequently, of IP-related applications also for businesses. In this context, many consider IP Communication as a killer application for the diffusion of broadband connections (OECD, 2006). "*IP Communication refers to the integration of data, voice and video solution into a single Internet Protocol (IP) - based network*". As a consequence, a large amount of different services and solutions are available for business application: advanced IP telephony calls and management, Web, audio-, video calls and videoconferencing, instant messaging, file sharing, calendar and other similar functions, email, fax, and voicemail, personal virtual assistance. An IP Communication-based network favours not only the exploitation, but also the convergence of different applications and in this way it allows an increase in services' quality and efficiency increase and, more than that, a decrease in network management costs.

The Advantages of IP Communication

IP Communication solutions have been available for some years. However, their diffusion among businesses and public organisations is still in its infancy, especially in countries that are late comers with reference to ICT adoption such as Italy. One of the most important reasons for this delay lies in the perception of the benefits deriving from IP communication. The conventional perception on convergence considers cost savings as the most compelling initial benefit of VoIP. A more progressive revenue path can be determined if companies start thinking about these applications according to their profitability, to the increase in efficiency or capability the new service introduces, to the maturity of the technology. In this respect, a common strategy based upon an infrastructure viewpoint is to first implement phone-to-phone services, then some value-added services. The resulting path starts from the development of mission critical services that add no value for the user and rely upon immature or developing platforms. A more suitable strategy would be to provide first a service that has value-added for the user, is not yet mission critical and is based on an emerging platform. As the technology matures, providers increasingly run existing services on the new platform (Corrocher, 2002).

In terms of cost reduction, IP Communication lowers the total cost of network management, by eliminating multiple sets of infrastructure, simplifying system administration and maintenance and consolidating voice and data circuits. Moreover, a single network permits to scale for the future and quickly react to the dynamically changing needs of the business. With reference to reduction in equipment and maintenance costs, we can observe that a converged network combines multiple network infrastructure into a single IP-based network, allows firms to centralize call processing, reduces the number of wiring drops in a new facility by 33% to 50% per user and lowers hardware connection costs, by reducing or even eliminating PBX/ACD upgrade costs. IP Communication also enables organisations to reduce network administration costs, by improving the productivity of network support staff through simplified network management, by minimizing outsourcing costs and by enabling network support staff to manage much larger user communities. Third, a converged network allows reducing network carrier costs.

If we turn to examine application-centric benefits, IP-based network enhances business communications, by providing a flexible infrastructure upon which all new applications and services can be easily deployed. These applications improve communications by facilitating increased mobility, delivering advanced functionality, and streamlining administrative tasks. As a result, employees are able to communicate more effectively and can focus on activities that create new revenue streams. With reference to this, IP Communication solutions are particularly beneficial for companies whose employees are often out of the office, since they remove geographical barriers and allow them to work from anywhere. Solutions like IP telephony, unified messaging, and IP contact centres have been designed to support today increasingly mobile workforce. For instance, unified messaging solutions deliver any type of message into a single inbox, giving employees the ability to access and manage their communications with different devices. Moreover, personal communication assistants help employees to be reachable anytime anywhere.

Problems and Open Issues in the Transition to IP Communication

Although the technology has been constantly improved, there remain some technical and regulatory problems for the provision of VoIP, which are mostly related to the characteristics of the IP network. These problems concern, on the one hand, quality of service (QoS), service reliability, and network migration; on the other hand they are related to the categorization of IP services within the general framework of communications services.

QoS might be undermined because packet-switched networks do not dedicate a path between the sender and the receiver. This means that each packet of information (voice, data or video) shares the available bandwidth with other packets. The packets sent through the Internet arrive at destination by a variety of sources and in any chronological order: they are accepted on a best-effort basis and can be delayed or lost. This implies that quality of service cannot be guaranteed. The more packets are lost, the more interpolation is needed and the more likely that the receiver of the call will hear distortion. Latency - the delay between when one party sends the voice/data message and the other hears/receives it - represents another important technical shortcoming (Howard et al., 1998). Other factors such as network ability to recover the original signals, quality of computing resources, voice processor and technical methods to establish connection between users are likely to affect the performance of IP communication solutions.

Reliability is also a critical concern for companies that are engaging in a network convergence process. By implementing the correct design, converged networks running VoIP can achieve a comparable level of reliability to the traditional network. The additional benefit of building redundancy into a converged network is that organizations can also improve the reliability of their data and video traffic. As far as the problem of migration to new networking solutions is concerned, the issue relates to companies' willing to protect existing investments, while migrating separate networks to a converged networking model. These investments include contractual agreements with service providers, which are generally long-term for local-government, education and healthcare organisations. In relation to this, most converged networking vendors have been developing products to ease this transition and ensure that new equipment can integrate with the existing infrastructure.

Finally, there is a lack of a clear international regulation about IP-based communications services. In 2003, the European Commission approved a general framework, which determines stronger restrictions for telecommunication incumbents than for independent providers. Currently, the main issues under investigation by regulatory authorities concern localization of users and emergency services, geographical numbers' provision, extra-territory service providers' presence, a standard quality of service and network reliability.

THE MAIN FOCUS: A FRAMEWORK FOR THE ANALYSIS OF THE ADOPTION OF IP COMMUNICATIONS

When studying individual choices of technology adoption, the empirical literature generally focuses on three sets of determinants: attributes of the technology, adopters' characteristics, and features of the competitive environment. Among adopters' characteristics, firm size, R&D expenditures, age, and capabilities are usually considered among the main determinants. The competitive environment can be described by the degree of firm concentration, the level of prices, and the existence of informational spillovers among potential users. Attributes of the innovation are usually related to technical features (i.e. compatibility with existing solutions) and may vary depending on the perception of potential adopters.

If we consider adoption of new technologies by public and private organisation, the decision to adopt is motivated by a series of factors such as the need of increasing operational efficiency, the willingness to improve market reach and profitability, and the need to manage risks (Daniel and Grimshaw, 2002). Several attributes of innovations act as drivers or obstacles of adoption. First and foremost, innovations must possess a relative advantage over the existing ideas. According to Rogers (1995), when a new technology emerges, individuals evaluate both its economic profitability and other variables - degree of risk, decrease in discomfort, savings in time and effort, immediacy of rewards. This concept is very similar to the idea of perceived usefulness of the technology, which identifies the subjective probability that using a specific technology would increase one individual job performance (Davis, 1989). Second, the extent to which an innovation is perceived as relatively difficult to understand and use - its complexity or perceived difficulty of use - can hinder the process of technology adoption. Complexity acts as an obstacle to the process of technology adoption, since implementing a new technology might require learning both at the individual and at the organisational level. The simplicity of a technology can be related to the capacity of different users to understand and use the technique without a great effort (Davis, 1989; Attewell, 1992; Mole et al., 2004). This idea recalls the notion of absorptive capacity as put forward by Cohen and Levinthal (1990), which identifies the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends. The absorptive capacity is a function of the firm prior related knowledge and characteristics of learning environment. Its development is highly path-dependent, since lack of investment in an area of expertise early on may hinder the future development of a technical capability in that area.

Potential users pay great attention also to the compatibility of an innovation, both in terms of technical features and in terms of past experiences and needs. Compatibility is especially important in the case of ICT, whose adoption is affected by the existence of network effects (Farrell and Saloner, 1985; Katz and Shapiro, 1985). In the presence of network effects, the utility from adoption increases in the number of other adopters that purchase the innovation. The trialability of innovations affects adoption since functioning, real world examples are often more important than arguments about advantages and expected functions. Experimenting with a new technology is an important benefit especially for early adopters, since they can only rely upon available information, while laggards can learn from other users' experience. Finally, an innovation is evaluated according to its observability, i.e. the extent to which its performance and related benefits are perceivable to the users and not only to the companies that produce it. Observability includes both "visibility" - the degree to which the innovation is visible - and "result demonstrability" – the degree to which the results of adopting the innovation are observable and can be communicated to others (Oh et al., 2003).

Turning to adopters' characteristics, one of the most relevant factors taken into account by the empirical literature is firms' size. This could be measured in many different ways (annual sales, total assets, total revenues and number of employees) depending both on the sector under investigation and on data availability. The role of size in the adoption of innovations is quite controversial, since the existing empirical investigations have produced mixed results. Many authors have found a positive relationship between size and timing of innovation adoption (Hannah and McDowell, 1984; Buzzacchi et al., 1995; Geroski, 2000; Hall-Khan, 2003). At the same time, studying 635 technological changes in the U.S. between 1970 and 1978, Acs and Audretsch (1988) found that small firms produce 2.5 times as many technological changes per employee as did larger firms. Furthermore, some authors (Pennings and Harianto, 1992; Buzzacchi et al., 1995) stress that the growing importance of external sources of technology could reduce the role of size as a determinant of adoption of new technologies. On the one hand, large firms can act as "first movers" in the process of ICT adoption because of the availability of professional competences and lower financial constraints; on the other hand, small firms can be less subject to legacy investments' constraints and adopt earlier. Other studies (Mohnen and Rosa, 1999; Daniel and Grimshaw, 2002) have underlined that firms' size might affect the perception of specific benefits and obstacles to innovation, with large firms more concerned about feasibility risks, high costs and uncertainty, and small firms hindered in the adoption by funding difficulties and lack of special equipment.

Geographical location can also be an explanatory factor of new technologies' adoption. The literature has emphasised that firms localized in the same area as other innovative organisations can more easily exploit knowledge spillovers (Acs et al., 1994). As far as ICT services and applications are concerned, several studies have highlighted the importance of location for the adoption/use of innovations (Prieger, 2003; Kim et al, 2006; Arbore and Ordanini, 2006).

Firms' performance, which is usually measured in terms of turnover, profits or value added, has also been investigated as a potential determinant of the adoption of new technologies. Although one could expect a positive relationship between profits and adoption of innovations, the literature does not usually find a significant relationship between the profitability of a firm and the adoption of new technologies. This is true for example for the banking sector, where the absence of such relation is interpreted as the lack of liquidity constraints on the technological strategies of banks (Hannan and McDowell, 1984).

Finally, the level of IT literacy and the previous experience in managing advanced technologies constitute important drivers for the propensity to adopt new technologies. These factors reflect the organisations' technological endowment in terms of inter-firm networking and the IT skills of firms' human resources. In order to take these variables into account, the literature has adopted different measures such as the presence of LAN, the access to broadband connections, the proportion of IT-skilled employees and the use of ICT services (Pennings and Harianto, 1992; Arduini et al., 2007, Corrocher and Fontana, 2007).

Data Collection and Explanatory Variables

To supplement the review of the technology carried out in the previous sections, we present now a study on the adoption of VoIP in Italy. The study is based on a dataset composed by 123 Italian organisations that have adopted IP Communication solutions from 2001, the year of introduction of IP Communication in Italy, to 2007. These organisations include hospitals and ASL (local healthcare organisations), universities and research centres, local government administrations and firms. Information was first collected through a series of interviews with managers and system engineers of Cisco Systems Italy, which is the worldwide market leader for networking solutions. This was crucial both to choose the possible organizations to be included in the analysis. Second, many complementary electronic resources were used to gain general information, balance sheets and income statements: AIDA, Bankscope and Isis[1] for private companies and financial institutions respectively; CINECA (a non profit Inter-University Consortium) and CNVSU (a National Committee for the Evaluation of the academic system) for universities and research centres; and the website of the Health Ministry and of single local government organisations (regions, provinces, and municipalities). Figure 1 reports the composition of our sample.

Almost 36% of the sample is made by firms: among these, 45% belong to manufacturing sectors, 32% to non financial services (e.g. consultancy, logistics), and 23% to the financial service sector. Concerning location, Table 1 below reports the joint frequencies of localization by type of activity.

The majority of companies (61.2%) are located in the North–West of Italy, especially around Milan, 2.3% in North–East, 27.3% in the Centre and 9.1% in the South. 23.6% of the sample is represented by education organisations and public administrations. Among universities and research centres, 41% is located in the South and 31%

Figure 1. Composition of the sample

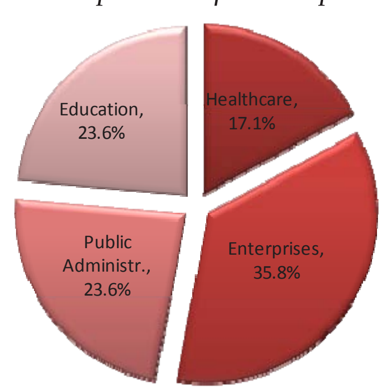


Table 1. Joint frequencies of type of organization by localisation

	North-West	North-East	Centre	South	Total
Healthcare	10.6%	4.1%	2.4%	0.0%	17.1%
Enterprises	22.0%	9.8%	0.8%	3.3%	35.8%
Public Administration	8.1%	3.3%	4.9%	7.3%	23.6%
Education	7.3%	4.1%	2.4%	9.8%	23.6%
Total	48.0%	21.2%	10.6%	20.3%	100%

in the North–West, while only few of them are located in the Centre (17%) or in the North–East (11%). These organisations on average adopted quite early. Public administrations include five regions, ten provinces, seven municipalities, and seven public companies. They are mostly concentrated in the North–West (34%) and in the South (31%) and generally adopted later than education organisations. Finally, healthcare organisations represent 17.1% of the sample, half of which are hospitals. There are no organizations in the South, and a few in the North-East (14.2%), while the Centre accounts for 23.8% and the North–West has 61% of the sample. Hospitals and ASL can be considered laggards since nobody has adopted before the end of 2005; however, they tend to acquire many IP telephones, especially because of their large size.

Intensity of adoption is reported in Figure 2. The lowest level of adoption is represented by just 1 IP phone while the maximum by 9250 phones. However, only four adopters have more than 4500 IP telephones. Not considering these outliers, the average intensity of adoption is represented by 387 IP phones. Looking at the distribution of this variable illustrated in the graph below, it is easy to see that there is a higher concentration of the sample around lower levels of intensity of adoption. Indeed, almost 40% of the organizations have adopted less than 100 IP phones and actually nearly 23% less than 50. Moreover, the concentration of the sample decreases as the intensity of adoption and the number of IPT acquired increases. Only 37.4% of the organisations have acquired more than 300 IP phones and 20.3% more than 700. This last group is mainly composed by the Public Sector, especially Hospitals and Universities, while enterprises are represented by Banks and Insurances.

Finally, we may look at the distribution of the sample in terms of size by intensity of adoption (Table 2).

Figure 2. Distribution of intensity of adoption

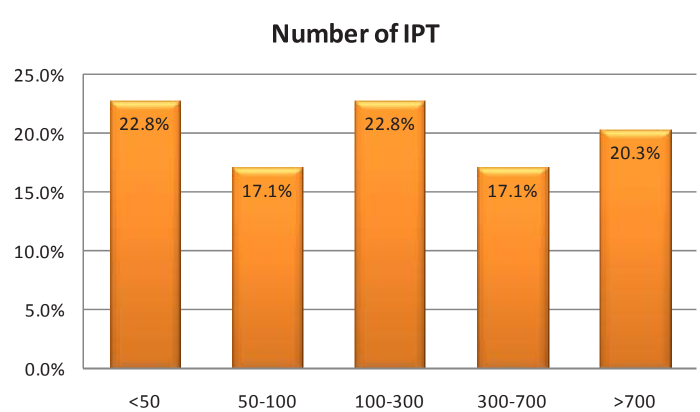


Table 2. Joint frequencies of size by intensity of adoption

	Size < 300	300 ≤ IPT < 1000	1000 ≤ IPT < 2500	2500 ≤ IPT < 4000	Size ≥ 4000	Total
IPT < 50	10.6%	5.7%	2.4%	3.3%	0.8%	22.8%
50 ≤ IPT < 100	7.3%	4.1%	1.6%	2.4%	1.6%	17.1%
100 ≤ IPT < 300	4.1%	8.9%	4.9%	3.3%	1.6%	22.8%
300 ≤ IPT < 700	0.8%	4.1%	8.9%	1.6%	1.6%	17.1%
IPT ≥ 700	0.0%	1.6%	4.1%	5.7%	8.9%	20.3%
Total	22.8%	24.4%	22.0%	16.3%	14.6%	100.0%

Frequencies in the Table highlight the presence of a positive relationship between these two variables. The larger an organisation is, the higher its level of adoption. Higher frequencies are found at the two opposite sides of the table. On the one hand, 10.6% of organisations with less than 300 employees have adopted less than 50 IP phones. On the other hand, 8.9% of organisations with more than 4000 employees have adopted more than 700 IP phones. In general, organisations in our sample tend to concentrate there in the upper-left and lower-right corners of the Table. At the same time, there are lowest frequencies for large institutions with minor intensity of adoption (0%, 1.6% and 3.3%) and by small organizations with major level of adoption (0%, 1.6% and 0.8%). Higher frequencies are located close to the diagonal. Thus, even medium sized organisations display an intermediate level of adoption intensity.

The Model and the Explanatory Variables

To investigate the determinants of VoIP adoption, we carry out a duration analysis. Our dependent variable is binary and takes the value 1 if adoption has occurred in a specific time period and 0 if not. An organisation becomes at risk of adopting after the VoIP has become available in Italy (i.e. January 2001). Time is measured in months since VoIP introduction. We estimate a logistic model with piece-wise constant hazard function. This specification allows splitting the adoption times at particular points (in our case quarters) so to ensure that events (i.e. adoptions) occur within each of the time intervals.

As far as covariates are concerned, we have constructed the following variables. First, we take into account the impact of Size, which is measured in terms of number of employees (in 000s). We expect size to be an important determinant of the diffusion of IP Communication. Many authors have found a positive relationship between size and innovation (Hannah and McDowell, 1984; Buzzacchi et al., 1995; Geroski, 2000; Hall-Khan, 2003). They demonstrate that large firms can act as “first movers” in the process of ICT adoption, since they can increase their image and exploit some advantages, such as relevant financial resources, best human capital and economies of scale and scope. However, other studies have shown that the relationship between size and adoption is much more complex with small and more dynamics firms being faster to adopt and the presence of curvilinear effects and ‘threshold effects’. To account for these possibilities, we also introduce Size^2 which is the square of the number of employees.

Second, we consider the role of investments. In order to measure these variables, we first homogenise private and public sector balance sheets and then create comparable indicators. Companies with negative CAPEX (expenditures on assets’ acquisition/upgrade) are not considered, since they do not invest in new assets, but rather disinvest their assets. Investments are defined as the total investments carried out in 2001 (Tot Invest. 01) the year VoIP became available. Since a unique measure of investments in innovation and particularly in ICT is not available, total investments (in 00000s) are used as a proxy. For private companies, we use CAPEX. For hospitals and ASL, we consider tangible assets and maintenance costs. For universities, we use expenditures in human resources and investment goods. Finally, for local government organisations, we use fixed investments. The level of investment is expected to positively influence the adoption of IP Communication.

Finally we consider the type of activity of the organisations. Our sample includes four macro categories: healthcare, public administrations, education and private companies. Healthcare includes both private and public hospitals and ASL; Education represents Universities and Research Centres; Public Administrations comprise Municipalities, Districts, and Regions, as well as few firms owned by local government organisations. The variable Public is a dummy equal to 1 if the organisation is public and 0 if not. We expect the type of activity to be an important determinant of adoption of IP Communication. In particular, there are important differences in managerial practices and behaviour between private and public sector. Private companies tend to be more inclined to adopt new technologies, as their structure allows them to take decisions more quickly and easily. Conversely, public organisations should be more reluctant and slow in the process of adoption, especially because of tenders' mechanisms that slow down the decision process.

Results

Table 3 reports our results. Covariates are introduced in sequence.

Specification (1) reports the baseline hazard and provides information on how the probability to adopt for the firms in our sample changes as time passes. The estimated coefficients on the duration time intervals are informative about the shape of the baseline hazard function. In particular, we can note that the hazard rises overtime. This result suggests that organisations become more likely to adopt as time passes and is robust to the inclusion of the other covariates in subsequent specifications.

In Column (2) we introduce our proxies for size. The coefficient of Size is negative and significant indicating that smaller firms have a relatively higher probability to adopt VoIP than large firms. This result is surprising since previous literature has found a positive relationship between size and the probability to adopt innovations given that large firms generally have more resources available. Our results seem to suggest instead that small firms are more dynamic and adopt earlier then bigger ones. Size^2 enters positively and significantly, albeit weakly, indicating that the relationship between size and the probability to adopt is indeed curvilinear. In particular, the probability to adopt decreases as the size of the firm increases up to a certain threshold after which it starts increasing. This result suggests that only very small firms and very big firms have a relative higher probability to adopt. Medium sized firms are those with the lowest probability to adopt.

In model (3) we introduce the variable total investment which captures the amount of resources spent by each organisation. This information is not available for all the organisations in our sample. Thus the number of observations in our sample drops from 972 to 499. The coefficient for this variable is positive and significant suggesting, as expected, that the probability to adopt is higher for those organisations that invest more. Interestingly, once we introduce our proxy for the amount of investment carried out by each organisation, our proxy for size loses significance. Indeed, there is a positive correlation between the size of the organisation and the amount of resources available for investment. It is also interesting to note that Size^2 remains weakly significant but it becomes negative thus indicating that when controlling

Table 3. Determinants of the probability to adopt VoIP for a sample of Italian Organizations

	(1)	(2)	(3)	(4)
$t < 4$	-5.731	-6.001	-7.703	-7.779
	[0.746]***	[0.766]***	[1.391]***	[1.392]***
$4 \leq t < 7$	-4.948	-5.214	-5.841	-5.911
	[0.688]***	[0.709]***	[1.049]***	[1.050]***
$7 \leq t < 10$	-3.930	-4.192	-4.782	-4.837
	[0.655]***	[0.677]***	[0.997]***	[0.996]***
$10 \leq t < 13$	-2.777	-2.996	-3.41	-3.466
	[0.649]***	[0.666]***	[0.961]***	[0.961]***
$t \geq 13$	1.846	2.341	2.71	2.448
	[0.621]***	[0.674]***	[0.984]***	[1.009]**
Size		-0.140	-0.011	0.126
		[0.071]**	[0.223]	[0.251]
$Size^2$		0.003	-0.059	-0.074
		[0.002]*	[0.037]*	[0.038]*
Tot Invest. 01			0.002	0.002
			[0.001]**	[0.001]**
Public				-0.474
				[0.428]
Observations	972	972	499	499
Log Likelihood	-251.53	-248.99	-114.80	-114.20
LR Chi-Square	149.72***	154.80***	95.26***	96.46***
Pseudo R-Square	0.229	0.237	0.293	0.296

*Logistic model with Piece-Wise Constant Hazard Function. Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.*

for the amount of investments, both small and large firms have a relatively lower probability to adopt and medium enterprises are those with the relatively higher probability to adopt. Finally, in model (4) we include the dummy PUBLIC. The coefficient of this dummy is not significant thus indicating that, after controlling for size and availability of financial resources, there is not statistically significant difference between private and public organisations in the probability to adopt VoIP.

All in all, our exploratory analysis of the determinants of adoption of VoIP in public and private organisations in Italy has suggested that two factors have driven the diffusion: the size of the organisation and the availability of resources to invest. Small firms seem to have reacted more quickly to the availability of the new technology and adopted faster then relatively larger ones, a result that suggest that rather than by congestion, adoption has been driven by information diffusion and hype about the new technology. However availability of financial resources is important too. Indeed, after controlling for the amount of investments carried out, size is no longer significant and organisation that invested more were relatively more likely to adopt.

CONCLUSION AND FUTURE TRENDS

This chapter has investigated the determinants of the adoption of IP communications solutions among public and private organisations in Italy. It has found, first, that organisations become more likely to adopt as time goes by, and second, that decision to adopt is mostly affected by size and availability of financial resources. Undoubtedly, an IP communication-based network allows exploiting a single infrastructure for all services – voice, data and video – and stimulating technological convergence. Organisations can benefit from this new technological system, because it offers cost savings and enhanced applications facilitating network management and on-line transactions. However, technical shortcomings, together with established habits and practices as well as legacy network investments can inhibit the adoption of IP communications. This might also explain why firms are more likely adopt as time passes and why small organisations are more inclined to adopt than larger ones.

The integration of different modes of communication is already happening and the increasing reliability of the network will foster the rate of adoption of VoIP. Organisations are implementing enhanced services over private networks and, if investments in bandwidth will help solve the problems of congestion and quality of services, companies will eventually run also voice over Intranets. The prospects for IP communication systems depend upon the interaction between demand and supply in the process of evaluating innovative solutions. On the one hand, technological developments – not only in the fixed but also in the mobile telecom environment - will play a crucial role in determining the opportunities for integrated applications. On the other hand, users' requirements and expectations will contribute to establish the future trajectory. In particular, efficiency advantages and the availability of integrated services can foster business users to move rapidly to Internet communication based upon intelligent terminals. A final consideration concerns the compatibility of the new technological solutions with users' managerial and organisational practices. In this respect, although the development of standards has enforced interoperability between hardware and software, there remain problems of compatibility between IP solutions and the existing modes of communication. For organisations that are already implementing integrated solutions on their networks, compatibility is mainly a technical issue; for late adopters who tend to be more inertial in putting new technologies into operation, compatibility represents an organisational issue.

REFERENCES

Acs, Z. J., Audretsch, D. B., Feldman, M. P. (1994). R&D Spillovers and Recipient Firm Size. *Review of Economics and Statistics*, *76*(2), 336-340.

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in Large and Small Firms: An Empirical Analysis. *The American Economic Review*, *78*(4), 678-690.

Arbore, A., & Ordanini A. (2006). Broadband divide among SMEs: The role of size, location and outsourcing strategies. *International Small Business Journal*, *24*, 83-99.

Arduini, D., Nascia, L., & Zanfei, A. (2007). A sectorial approach to the diffusion of ICT: Empirical evidence on Italian firms. Working paper presented at the DRUID Summer Conference 2007.

Attewell, P. (1992). Technology Diffusion and Organizational Learning: The Case of Business Computing. *Organization Science*, *3*(1), 1-19.

Babbage, R., Moffat, I., O'Neill, A., & Sivaraj, S. (1997). Internet phone: Changing the telephony paradigm? *BT Technology Journal*, *15*(2), 145-157.

Buzzacchi, L., Colombo, M., & Mariotti, S. (1995). Technological regimes and innovation in services: the case of the Italian banking industry. *Research Policy, 24,* 151-168.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly, 35,* 128-152.

Corrocher, N. (2002). The diffusion of Internet telephony among consumers and firms: Current issues and future prospects. *Technological Forecasting and Social Change, 70,* 525-544.

Corrocher, N., & Fontana, R. (2007). Expectations, Network Effects and Timing of Technology Adoption: Some Empirical Evidence from a Sample of SMEs in Italy. forthcoming in *Small Business of Economics.*

Colombo, M. G., & Grilli, L. (2006). *Broadband access and broadband-based application: an empirical study on the determinants of adoption among Italian SMEs.* Working paper. Politecnico di Milano.

Daniel, E. M., & Grimshaw, D. J. (2002). An exploratory comparison of electronic commerce adoption in large and small enterprises. *Journal of Information Technology, 17,* 133-147.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319-340.

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *RAND Journal of Economics, 16*(1), 70-83.

Hannan, T. H., & McDowell, J. M. (1984). The determinants of technology adoption: the case of the banking firm. *RAND Journal of Economics, 15,* 328-335.

Hall, B. H., & Khan, B. (2003). *Adoption of New Technology.* Working paper. Institute of Business and Economic Research, University of California, Berkley.

Howard, J., Intven, H., & Zohar, M. (1998), *Internet Telephony – The Regulatory Issues.* McCarthy Tetrault Publications.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition and compatibility. *American Economic Review, 75*(3), 424-440.

Kim, D., Cavusgil, S. T., & Calantone, R. J. (2006). Information system innovations and supply chain management: channel relationships and firm performance. *Journal of the Academy of Marketing Science, 34*(1), 40-54.

Krupinski, D. (1997), *Introduction to Internet Telephony.* http://www.phonezone.com/tutorial/it-intro.htm.

Mole, K. F., Ghobadian, A., O'Reagan, N., & Liu, J. (2004). The Use and Deployment of Soft Process Technologies within UK Manufacturing SMEs: An empirical Assessment Using Logit Models. *Journal of Small Business Management, 42*(3), 303-324.

OECD (2006), *VoIP: Developments in the market.* Working Party on Telecommunication and Information Services Policies, OECD, Paris.

Oh, S., Ahn, J., & B. Kim (2003). Adoption of broadband Internet in Korea: the role of experience in building attitudes. *Journal of Information Technology, 18,* 267-280.

Ono, R., & Aoki, K. (1998). Convergence and new regulatory frameworks. *Telecommunication Policy, 22*(11), 817-838.

Pennings, J. M., & Harianto, F. (1992). The diffusion of technological innovations in the commercial banking industry. *Strategic Management Journal, 13,* 29-46.

Prieger, J. E. (2003). The supply side of the digital divide: Is there equal availability in the broadband Internet access market? *Economic Inquiry, 41*(2), 346–363.

Rogers, E. M. (1995). *Diffusion of Innovations.* New York: Free Press.

Saloner, G., & Shepard, S. (1995). Adoption of Technologies with Network Effects: an Empirical Examination of the Adoption of Automated Teller Machines. *RAND Journal of Economics, 26*, 479-501.

KEY TERMS

Automatic Call Distribution (ACD): In telephony, an Automatic Call Distributor (ACD) is a device or system that distributes incoming calls to a specific group of terminals that agents use. It is often part of a computer telephony integration system.

Information and Communication Technology (ICT): Information and Communications Technology (ICT) is a term used to indicate a broad subject concerned with technology and other aspects of managing and processing information.

Internet Protocol (IP): The Internet Protocol (IP) is a data-oriented protocol used for communicating data across a packet-switched internetwork.

Private Branch Exchange (PBX): A PBX is a premises switching system, serving a commercial or government organization, and usually located on that organization's premises. PBXs provide telecommunications services on the premises or campus, (e.g., internal calling and other services), and access to public and private telecommunications network services.

Protocol Independent Multicast (PIM): Protocol-Independent Multicast (PIM) is a family of multicast routing protocols that can provide one-to-many and many-to-many distribution of data over the Internet. The "protocol-independent" part refers to the fact that PIM does not include its own topology discovery mechanism, but instead uses routing information supplied by other traditional routing protocols.

Public Switched Telephone Network (PSTN): A Public Switched Telephone Network (PSTN) denotes those portions of the local exchange carrier networks that provide public switched telephone network services.

Voice over Internet Protocol (VoIP): Voice over Internet Protocol (VoIP) is a protocol optimized for the transmission of voice through the Internet or other packet switched networks. VoIP is often used abstractly to refer to the actual transmission of voice (rather than the protocol implementing it).

Chapter XXIX
Intelligent Networking and Business Process Innovation:
A Case Study Analysis of Home Box Office and Dell Computers

Richard A. Gershon
Western Michigan University, USA

ABSTRACT

Today, innovation is much more about much than just developing new products. It is about reinventing business processes and building entirely new markets to meet untapped customer needs. This chapter will examine the subject of business process innovation which involves creating systems and methods for improving organizational performance. Special attention is given to the topic of intelligent networking which represents the combination of software, technology, and electronic pathways that makes business process innovation possible for both large and small organizations alike. A central tenet is that the intelligent network is not one network, but a series of networks designed to enhance world-wide communication for business and residential users. Two very different kinds of intelligent networks are discussed in this chapter. The first involves satellite-to-cable television networking where the emphasis is on program distribution to the end consumer. The second is a supply chain management network where the emphasis is on just-in-time manufacturing. Each of the said networks represents a highly innovative business process and share the common goal of improving organizational performance. The information presented in this chapter is theory-based and supported by a case-study analysis of Home Box Office, Inc. and Dell Computers.

INTRODUCTION

International business has been transformed by the power of instantaneous communication. The combination of computer and telecommunications have collapsed the time and distance factors that once separated nations, people and business organizations (Friedman, 2005). The information economy involves the full integration of transnational business, nation-states and technologies operating at high speed. It is a global economy that is being driven by free-market capitalism. The basic requirements for all would be players is free trade and a willingness to compete on an international basis (Friedman, 2005). The once highly centralized business has given way to the transnational corporation (TNC) that operates in multiple countries throughout the world. Instead of time and communication being highly synchronized, today's working professional lives in a digital world of asynchronous and virtual communication that allows for the international collaboration of projects regardless of time zones, geographical borders and physical space (Gershon, 2002). The driving force behind such changes is that today's information economy is built on what Noam (2001) calls a "network of networks." (pp. 1-2). We don't talk to people, we network with them.

The Intelligent Network

The intelligent network represents the combination of software, technology and electronic pathways that makes business process innovation possible for both large and small organizations alike. The intelligent network can be likened to the internal nervous system of an organization. It provides the basis for the seamless integration of information and communication both internal and external to the organization. The combination of high-speed voice, data and video communication allows today's business enterprise the ability to coordinate the production, marketing and delivery of products on a worldwide basis.

The Foundation Principles. A central argument of this chapter is that the intelligent network is not one network, but a series of networks designed to enhance worldwide communication for business and residential users (Noam 2001). What gives the network its unique intelligence are the people and users of the system and the value-added contributions they bring to the system via critical gateway points. Systems theorists refer to this as a the principle of network wholism (Gershon, 2004). As an example, the Internet has become greater than the sum of its parts. The exponential growth of the Internet is due to the many contributions of its users in terms of search engine design (*Google*); electronic commerce, (*eBay, Amazon.com* etc.); social networking; (*Facebook, My Space* etc.); speed and throughput (cable modems, DSL etc.) Today, the Internet has created a new business model that maximizes the potential for instantaneous communication to a worldwide customer base. It has fundamentally changed how retailtrade is conducted in terms of information gathering, marketing, production and distribution.

Intelligent Networking and Organizational Decision Making Process. The intelligent network has had a major effect on the spatial reorganization for today's highly complex organization. Time and distance factors have become less important in determining where a company chooses to locate (O'Hara-Devereaux, & Johansen, 1994; Poole, 1990). One important consequence is that organizational hierarchies tend to beflatter, thereby, allowing direct communication between and among organizational players (Huber, 1990). What's important to remember, is that intelligent networks are not stand alone entities. Rather, intelligent networks are part of a greater human and organizational decision-making process (Monge & Contactor, 2003; Monge & Eisenberg, 1987). Technology alone is rarely the key to unlocking economic value. Companies create real wealth when they combine technology with new ways of doing business (Manyika, Roberts & Sprague, 2007). In sum, people and innovative thinking make the difference.

What is Innovation?

Innovation can mean the introduction of an entirely new product or service to the market-place. But innovation can also mean introducing an entirely new process that improves long term organizational performance. In the field of business and technology, innovation breaks down into three subset areas. They include: 1) product innovation, 2) business model innovation and 3) process innovation. Some of today's more innovative companies are innovative in all three areas. The combination of the Apple iPod and iTunes media store, for example, qualify as a new product (i.e., *iPod*), business model (*iTunes*) and business process (*iTunes + MP3 music file sharing delivery*). In sum, Apple has created the first sustainable on-line music delivery business of its kind. Innovation is important because it creates a lasting advantage to a company or organization.

This chapter will look at the relationship between intelligent networking and business process innovation. Two very different kinds of intelligent networks will be considered in this examination. The first kind of network looks at the principles of cable networking with a specific reference to satellite-to-cable television distribution where the emphasis is on achieving economies of scale in design and operation. The second kind of network examines a supply chain management system network where the emphasis is on global inventory management and just-in-time manufacturing. Each of the said networks represent a highly innovative business process and share the common goal of improving organizational performance. The information presented in this chapter is theory-based and supported by a case-study analysis of Home Box Office, Inc. and Dell Computers.

BUSINESS PROCESS INNOVATION

Today, innovation is about much more than developing new products. It is about reinventing business processes and building entirely new markets to meet untapped customer needs. Davenport & Short (1990) define *business process* as "a set of logically related tasks performed to achieve a defined business outcome." A process is a structured, measured set of activities designed to produce a specified output for a particular customer or market. It implies a strong emphasis on how work is done within an organization (Davenport, 1993). In their view, a business process exhibit two important characteristics: 1) they have customers (internal and external) and 2) they cross organizational boundaries, (i.e., they cut across different organizational subunits). One technique for identifying business process in an organization is based on the principles of value chain analysis proposed by Porter (1985). Value chain analysis includes the various activities through which a firm develops a competitive advantageand creates shareholder value. The goal is to offer the customer a level of value that exceeds the cost of the activities, thereby, resulting in a profit margin.

Business process innovation involves creating systems and methods for improving organizational performance. The application of business process innovation can be found in a variety of settings and locations within an organizational structure, including product development, manufacturing, inventory management, customer service, distribution etc. A highly successful business process renders two important consequences. First, a highly successful business process is transformative; that is, it creates internal and external efficiencies that provides added value to the company and organization. Second, it sets into motion a host of imitators who see the inherent value in applying the same business process to their own organization. Table 1. provides a comparison of five media and telecommunications companies that are industry leaders in the use of business

Table 1. Five media and telecommunications companies and the transformative impact of business process innovation

Home Box Office	In 1975, helped advance the principle of satellite/ cable networking by using satellite communication to advance long haul television distribution.
Dell Computers	In the area of computer manufacturing, Dell created a highly successful business model utilizing just-in-time manufacturing techniques as well as direct-to-home sales capability.
Pixar Studios	Developed computer generated animation graphics in contrast to traditional cartoon animation techniques. Examples include, *Toy Story, Finding Nemo, Monsters Inc.,The Incredibles, Cars* etc
Apple Computer	The combination of the Apple iPod and iTunes media store have created the first sustainable music down-loading business model of its kind.
Netflix	Has become the largest on-line DVD rental service in the U.S., offering flat rate rental by-mail to customers. Developed highly sophisticated supply chain management system.

process innovation. Each of the said companies have rendered a host of imitators in the way they have refined business process.

BUSINESS PROCESS AND REENGINEERING

Business process innovation involves creating systems and methods for improving organizational performance. Specifically, business process innovation is about creating internal and external efficiencies that will improve operational design and performance. Accordingly, process innovators are obsessive problem solvers. There is a constant focus on finding new ways to improve quality and optimize performance. Everyone within the organization from senior management to the worker on the floor has a responsibility to improve product and service quality.

Reengineering

The principles of reengineering owe their aegis to an evolving set of business process improvement theories, starting with Total Quality Management (TQM) and including Six Sigma. While a number of management scholars have contributed to the principles of reengineering, Hammer & Champy (1993) popularized the term to the general public when they published *Reengineering the Corporation: A Manifesto for Business Revolution*. Since then, the term "reengineering" has become part of the corporate lexicon whenever an organization should decide to reorganize or downsize its business operations. Reengineering, however, is really about business process innovation.

The decision to reengineer usually comes about at a time when a business is faced with major competitive threats or recognizes that its operations are costly and inefficient. Another term for reengineering is *business process redesign*. Hammer & Champy state the question very simply. "If I was recreating this company today, given what I know and given current technology, what would it look like?" (p. 31). Business process redesign often means starting over. Reengineering represents a fundamental rethinking of business processes in order to bring about dramatic improvements in organizational performance. It means throwing out old assumptions about how things were done in the past and developing new procedures and solutions.

In order to accomplish such goals, reengineering means having to rethink key business processes and a willingness to abandon old or outmoded ways of doing business. Improvements in performance can be measured in several ways, including reduced manufacturing costs, greater speed and efficiency, improved customer service etc. While the specifics of reengineering will vary from one organization to the next, there are certain features that are typical of a reengineered process. This includes:

- Creating cross-functional teams
- Streamlining the business process
- Designing multiple versions of a business process
- Sharing information and resources

One of the basic tenants of reengineering is that one organizes around keybusiness process which may be handled by a small, cross functional work team rather than by rigidly defined organizational hierarchies, departments or assembly lines. A cross functional team consists of members from various departments with different skill sets. They meet regularly as a group to solve ongoing problems of mutual interest. Second, reengineering presupposes the ability to identify customer needs and then designing a process and aligning people to meet those needs. A routine request for information, for example, should not be routed among five different departments. Instead, the request should be handled by one person who is given the proper resources and authority to handle such requests. That person now performs the whole process and also serves as the single point of contact for the customer.

In years past, large organizations tended to compartmentalize information. It was not uncommon to find that several divisions within an organization might create their own separate data base. The duplication of effort was both costly and inefficient. Today, the emphasis is on the sharing of information resources across divisional lines thus promoting greater efficiency in product manufacturing, marketing and distribution. This is at the heart of Enterprise Resource Planning (ERP) which allows various players within a company's extended supply chain to access the same information from a commonly agreed upon information system platform (Zheng, Yen & Tarn, 2000).

HOME BOX OFFICE, INC.

The real move to modern cable television began on November 8, 1972, when a fledgling company named Home Box Office (HBO) began supplying movies to 365 subscribers on the Service Electric Cable TV system in Wilkes Barre, Pennsylvania. That night, Jerry Levin, then Vice-President for Programming, introduced viewers to the debut of HBO. The feature programming for that inaugural night was a hockey game between New York and Vancouver and a film prophetically entitled, *Sometimes a Great Notion*.

From the beginning, HBO developed two important innovations that helped to promote its rapid growth and development. First, HBO introduced the principle of premium television (i.e., business model innovation). Specifically, HBO achieved what no other television service provider had accomplished to date; namely, getting peopleto pay for television. The principle of advertiser supported "free" television was firmly engrained in the mind's of the American public. What HBO did was change public perception about the nature of television entertainment. HBO offered a unique value proposition emphasizing recently released movies and other specialized entertainment that could not be found elsewhere on the general airwaves. While HBO was not the first company to introduce a monthly per channel fee service, they were the first to make it work successfully. This marked the beginning of premium television entertainment (Parsons, 2003). Second, HBO utilized microwave and later satellite communications for the transmission of programming, rather than distribution by videotape (process innovation). Prior to HBO, there was no precedent for the extensiveuse of satellite delivered programming in the U.S.

Satellite/Cable Networking as a Business Process

A communication satellite is essentially a microwave relay in the sky, operating at 22,300 miles above the earth's equator. It receives microwave signals in a given frequency and retransmits them at a different frequency. Satellites provide an efficient means of reaching isolated places on the earth and are considerably less expensive than terrestrial communication links for select applications. Communication satellites are a versatile form of wireless communication. What distinguishes communication satellites from other forms of wireless communication is its high orbital position and movement.

The term "geosynchronous orbit" refers to a satellite that operates at 22,300 miles above the earth's equator. The satellite rotates at the speed of the earth. Hence, the satellite appears to be stationary in its orbital position. In principle, a complete satellite link requires a line of sight path extending between the earth station and the satellite. A satellite footprint refers to the signal's area of coverage. The HBO signal, for example, utilizes an east Coast and west Coast feed in order to blanket the entire U.S. Therefore, any cable operating system equipped with an earth station that falls within the footprint of a satellite fed signal and that is locked on to the appropriate transponder is capable of receiving the same signal.

Signal Quality and Strength. In designing the proper satellite communication link, there comes a tradeoff in design. Either one builds small satellites with large earth stations or large satellites with small earth stations. In short, amplification of the signal has to happen at one end or the other. During the decades of the 1960's and 1970's, satellite manufacturers like RCA and Ford Aerospace built small satellites and large earth stations because there was no practical way to launch large satellites with heavy payloads. Subsequent improvements in satellite design and launch capability have allowed designers to put more amplifying power in the satellite or space segment end. This, in turn, has led to a corresponding decrease in the size of earth stations. The direct beneficiaries of such design changes have been the broadcast, cable and direct broadcast satellite television industries (Gershon, 2000; Parsons & Frieden, 1998).

Satellite Advantages

1. ***Economies of Scale.*** Cost bears no relationship to the distance involved and/or to the number of users. When considering any distance greater then a few hundred miles, the cost of broadcasting via satellite is significantly less expensive than landline transmission. This is because only one relay station is involved; namely, the satellite. The satellite's footprint (or area of coverage) permits many earth stations to simultaneously receive the same signal. Therefore, an economy of scale is realized because it costs no more to transmit television to one earth station then it does 10,000 so long as they fall within the same footprint. This is the key technical / economic underpinning that makes cable television and direct broadcast satellite television possible.
2. ***Wide Area of Coverage.*** Satellites provide wide area coverage where distance and terrain are not critical factors. This becomes especially important for mobile communication for ships at sea or fleet management. Satellite communication is also good in rural areas that may not have established terrestrial communication links. This is one of the special appeals of Direct Broadcast Satellites for people living in the rural countryside.

3. ***Rapid Installation of the Ground Network.*** The installation of an earth station is relatively inexpensive and can be accomplished quickly. In addition, there is a lower cost per added site. This has been especially important for mobile news operations in the field as well as military communication.

HBO and the Principle of Cable Networking

HBO's 1975 decision to use satellite communications was significant in two ways. First, it demonstrated the feasibility of using satellite communication for long haul television distribution. As a consequence, HBO was able to create an efficient distribution network for the delivery of its programming to cable operators. Second, the development of the satellite/cable interface would usher in a whole new era of cable programmers that were equally capable of leasing satellite time and delivering their programs directly to cable operating systems, including: WTBS, 1976; ESPN, 1979; CNN, 1980; and MTV, 1981. Thus, was born the principle of cable networking; that is, television programming designed exclusively for cable operating systems and later direct broadcast satellite systems (Gershon & Wirth, 1993). The principle of satellite / cable networking would transform the business process of long-haul television distribution. As cable analyst, Paul Kagan once remarked:

Rarely does a simple business decision by one company affect so many. In deciding to gamble on the leasing of satellite TV channels, Time Inc. took the one catalytic step needed for the creation of a new television network designed to provide pay TV programs. (HBO, Inc., 1984).

Today, HBO has extended its brand worldwide and reaches an estimated 18 million subscribers in more than 50 countries in Latin America, Asia and Central Europe.

DELL COMPUTERS

The company known as Dell Computers was established by Michael Dell in 1984 and has grown to become one of the world's preeminent manufacturers of desktop and laptop computers. Dell builds computers to customer order and specification using just-in-time manufacturing techniques. The company has built its reputation on direct sales delivery to the end consumer combined with strong customer support. Dell's business model is simple in concept, but very difficult to execute in practice (Kraemer & Dedrick, 2002).

Michael Dell started out as a pre-med student at the University of Texas. Dell soon became fascinated by computers and created a small niche in the assembly and sale of PCs and PC components out of his dormitory room. Dell bought excess supplies at cost from IBM dealers which allowed him to resell the components at 10-15 % below the regular retail price. He then began to assemble and sell PC clones by purchasing retailers' surplus stock at cost and then upgrading the units with video cards, hard disks, and memory. Dell then sold the newly assembled IBM clones at 40% below the cost of an IBM PC (Thompson & Strickland, 2006). By April 1984, with sales reaching $80,000 a month, Dell dropped out of the university and formed a company called PCs Limited. The ability to sell directly to the end user at a discounted price proved to be a winning formula and by the end of 1986, sales had reached $33 million. PCs Limited was renamed Dell Computers in 1987 and the company soon opened its first set of international offices.

From 1990 to 1993, Dell experimented with traditional retail distribution in hope of faster growth, but soon realized that such methods were less profitable and refocused on direct sales. By 1996, Internet sales had taken off and the company realized that computer savvy shoppers preferred the convenience of custom ordering what they wanted directly from Dell and having it delivered to their door. During this time, Dell had become master innovators involving two important business processes. The first process was customization using a just-in-time manufacturing capability. Dell built computers to customer order and specification, thereby, eliminating excess inventory and the need for storage. The second important process was direct-to-consumer sales delivery thus avoiding costly investment in retail store infrastructure. It was process model that other computer manufacturers would later adopt (Fields, 2006).

Supply Chain Management as a Business Process

Supply chain management (SCM) is a complex business model that takes into consideration the entire set of linking steps necessary to produce and deliver a product to the end consumer. A supply chain consists of the intermediary steps from the point of product inception (inclusive of the purchase order) to the delivery of the finished good(s) to the consumer. A supply chains consists of the following intermediary steps:

- Research and design teams
- Raw Material Suppliers
- Intermediate Product Manufacturing
- End Product Manufacturing
- Wholesalers
- Retailers and/or Direct Sales

SCM has two distinct and equally important parts: 1) the philosophy and 2) the methodology. SCM philosophy is grounded in the belief that everyone involved in the supply chain is both a supplier and customer and requires access to timely, up-to-date information. The goal is to optimize organizational efficiency and to meet the needs of any and all suppliers and customers. SCM methodology has to do with the specifics of strategy implementation. Through process reengineering, all non-essential elements are eliminated. SCM forces companies to move away from an organizational structure designed around functional silos toward one designed around the end-to-end flow of business processes. Information is key. To that end, an essential element of any SCM methodology is the ability to share timely information across the entire supply chain system. A well designed SCM system gives automated intelligence to an extended network of suppliers, manufacturers, distributors, retailers and a host of other trading partners (Tarn et.al., 2002; Zheng, Yen & Tarn 2000).

Enterprise Resource Planning. A supply chain is connected by transportation and storage activities and coordinated through planning and networked information activities. When engineers discuss the architecture of a network; they are describing how the physical parts of the network are organized, including: Information pathways (configurations), 2) Terminals (gateways and access points) and 3) Data enhancement equipment (software protocols and add-on devices). Central to any discussion of supply chain management and intelligent networking is the principle of enterprise resource planning (ERP) which attempts to integrate all departments and functions across an entire company onto a single computer system using a common data base and a shared set of reporting tools (Tarn et al., 2002). Dredden & Bergdolt (2007) define enterprise resource planning as "information systems that integrate processes in an organization using a common database and shared reporting tools." (p. 48).

Just-in-Time Manufacturing

Telecommunications has collapsed the time and distance factors that once separated nations, people and business organizations. Communication is instantaneous. The combination of high-speed voice, data and video communication allows both large and small organizations the ability to coordinate the production, marketing and delivery of products on a worldwide basis. The full impact of instantaneous communication can be seen in the area of SCM, global inventory management systems and just-in-timemanufacturing capability.

Most companies have access to excellent hardware and software capability that enables them to operate in an international business environment. The distinguishing factor often centers on speed and turn around time. Faster product cycles and the ability to train and produce worldwide production teams have transnationalized the manufacturing and distribution process. It is the ability to apply time-based competitive strategies at the global level that enables the transnational corporation to manage inventories across borders. At the heart of time base competitiveness is just-in-time manufacturing which allows a company to meet an order in the least amount of time. Just-in-time manufacturing (and delivery) relies on the use of supply chain management and ERP systems for the purpose of tracking customer orders. ERP systems are designed to interface with Universal Product Codes (i.e., bar codes) or Radio Frequency Identification (RFID) tags which enables a manufacturer or service provider to track the status of a product throughout theentire manufacturing and delivery cycle. In sum, SCM systems integrate and optimize both internal and external processes to the organization. In contrast, ERP systems tend to focus on internal business processes within the boundaries of a single organization. The ERP system coordinates and integrates all information planning activities within a single organization (Tarn et. al., 2002). This can include reacting to customer needs (i.e., answering customer inquiries about production status, delivery dates etc.)

Dell and Global Supply Chain Management

Today, Dell has an international workforce of 35,000 employees located in 34 countries and three major regions of the world, including the Americas, Europe/the Middle East and Asia Pacific. Dell's selection of geographic location and production facilities has largely been driven by it foreign direct investment strategy, including the perceived profitability of the market and growth potential. Each of the three regional hub sites have their own headquarters and set of assembly plants. Because of Dell's build-to-order philosophy, Dell has evolved a highly sophisticated manufacturing and logistics capability. A global network of suppliers and contract manufacturers support each production facility. Instead of producing all the necessary components itself, Dell contracts with other manufacturers to produce subassembly parts, such as motherboards, microprocessors, monitors etc. Dell maintains control over the final assembly portion, paying particular attention to customized feature elements (Kraemer & Dedrick, 2002)..

Dell's global inventory management system requires an efficient method ofcommunication in order to meet customer demands and to ensure a ready supply of parts on hand to support various kinds of configuration requests. Over time, Dell has builta complex, global wide SCM / ERP system that tracks information between and among suppliers, distributors, and other key component players that involve product manufacturing and support. In addition, Dell has established a specific network of suppliers and contract manufacturers to support each production facility.

In the past, Dell's approach to computer manufacturing involved a standardized assembly line process, whereby, a single individual would install a single component and the partly assembled PC was sent on to the next station. In 1997, Dell undertook a major process redesign known as "cell manufacturing" by which a team of workers would work together to assemble an entire PC at a workstation or cell (Thompson & Strickland, 2006). This technique has resulted in a steep decline in assembly time and increased productivity per square foot of assembly space. According to CEO Michael Dell, this has allowed us "to drive for even greater excellence in quality, cycle time and delivered cost. We will innovate and adapt our supply chain model to help drive differentiated product design, manufacturing and distribution models." (Hoffman, 2007).

DISCUSSION

There are no short cuts when it comes to innovation. As Hoff (2004) notes, "inspiration is fine, but above all, innovation is really a management process." (p. 194). Putting the right structures people and processes in place should occur as a matter of course – not as an exception (McGregor, 2007). Accordingly, business process innovation (and the intelligent networks that support it) are part and parcel of a greater human and organizational decision-making process (Monge & Contractor, 2003). The goal is to improve organizational performance. HBO's pay cable television service was greatly enhanced by the first-of-its kind satellite-to-cable distribution system thus advancing the cause of cable networking. Dell computer's direct to home retail sales strategy was greatly aided by its worldwide supply chain management system combined with its just-in-time manufacturing capability. Both companies are major innovators in the area of improving organizational performance. This can be seen in Table 2.

Value Creation

What is the value of one good idea or suggestion? Business process innovation is about creating a special value to the organization. Innovation, without value creation, is simply a technology driven effort that may provide incremental improvements to the organization, but does not address the larger question of how to make the organization better. Value creation can translate in many ways and formats. Not every innovative solution has to be a major breakthrough. Sometimes, small incremental changes in the area of business process can make a big difference to an organization in terms of product quality, production and distribution efficiency, cost containment and/or customer service. The Japanese auto industry use the term *kaizen* to describe the principle of continuous improvement. Dell applied the very principle when the company improved its approach to computer manufacturing (and customization) by introducingits cell manufacturing technique. This, in turn, has added value to the company's ongoing just-in-time manufacturing capability.

Developing a Culture of Innovation

Companies, like people, can become easily satisfied with organizational routines that stand in the

Table 2. Comparison of business process innovation and application

	Dell Computers	**HBO**
Major Business Areas Supported	The manufacture of desk top and lap top computers.	The production and delivery of pay cable television services.
Planning Goals	To support extended supply chain management system and help advance enterprise resource planning.	To support cost-effective delivery of television programming to U.S. cable operating systems.
Major Functionality and Benefits	Optimize information flow throughout the extended supply chain; thus enabling global inventory management and just-in-time manufacturing capability.	Satellite-to-cable network interface creates economies-of-scale cost savings for television production company.
Relationship with Customers	Provides timely information on production status and makes possible direct-to home delivery of computer equipment.	Provides premium television services to the end consumer via cable television or DBS.

way of being innovative. Managers can sometimes become preoccupied with fine-tuning and making slight adjustments to an existing product line or business process rather than preparing for the future. Forward thinking companies must be able to deconstruct management orthodoxy. HBO demonstrated this willingness by changing the public perception about the nature of television entertainment. HBO effectively created a new value proposition that television was something worth paying for; especially if the company could deliver premium quality television that was otherwise unavailable to the consumer in 1975. Hence the company's future slogan,it's not just television – it's HBO. But equally important, HBO changed the business process of how television programming was to be delivered to the cable operator, who effectively served as the company's retail outlet. By introducing the satellite / cable interface, HBO changed the business process of long-haul television distribution forever more. As Pilotta, Widman & Jasko (1988) point out, organizations (even large ones) are always human constructions; that is, they are made and transformed by individuals. Strong, innovative companies succeed by creating a culture where everyone has a role to play in making the organization better.

REFERENCES

Davenport, T., & Short, J. (Summer, 1990). The new industrial engineering: Information technology and business process redesign. *Sloan Management Review*, 11-27.

Davenport, T. (1993). *Process innovation*. Boston, MA. Harvard Business School Press.

Dredden, G., & Bergdolt, J. (2007). Enterprise resource planning. *Air Force Journal of Logistics*, *2*(31), 48-52.

Fields, G. (2006). Innovation, time and territory: Space and the business organization of Dell Computer. *Economic Geography*, *82*(2), 119-147.

Friedman, T. (2005). *The world is flat*. New York, NY: Farrar, Straus & Giroux.

Gershon, R. A. (2004). Intelligent networking and the information economy. In K. Lewandowski (Ed.), *Annual Review of Communications*, *57*, 611-622. Chicago, IL: International Engineering Consortium.

Gershon, R. (2002). The Transnational media corporation and the economics of global competition. In Yahya R. Kamalipour (Ed.), *Global Communication* (pp.51-73). Belmont CA: Wadsworth.

Gershon, R. (2000). *Telecommunications management: Industry structures and planning strategies.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gershon, R., & Wirth, M. (1993). Home Box Office: The emergence of pay cable television. In R. Picard (Ed.), *The Cable Networks Handbook* (pp. 114-122). Riverside, CA: Carpelan Press.

Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*. New York, NY: Harper Business.

Hammer, M., & Stanton, S. (1995). *The Reengineering Revolution: A Handbook***.** New York, NY: Harper Business.

HBO (1984). *Pay TV guide: Editor's pay TV handbook*. New York: NY

Hoff, R. (2004, October 11). Building an idea factory. *Business Week* (p. 194).

Hoffman, W. (2007, February 26). Dell's logistics restart. *Traffic World* (p. 1).

Huber, G. (1990). A theory of the effects of advanced information technologies on organizational design, intelligence and decisionmaking. *Academy of Management Review*, *15*, 195-204.

Kraemer, K., & Dedrick, J. (2002). Dell computer: Organization of a global production network." Available at: http://www.crito.uci.edu/GIT/publications/pdf/dell.pdf Retrieved: 6 October 2006.

Manyika, R., Roberts, R., & Sprague, K. (2007, December). *McKinsey Quarterly*. Available at: http://yaleglobal.yale.edu/display.article?id=10075. Retrieved, 15 January 2008.

McGregor, J. (2007, May 14). *Most innovative companies* (p. 60).

Monge, P., & Eisenberg, E. (1987). Emergent communication networks. In F. Jablin, L. Putnam, K. Roberts, & C. O'Reilly (Eds.), *Handbook of Organizational Communication* (pp. 304-342). Norwood, NJ: Ablex.

Monge, P., & Contractor, N. (2003). *Theories of Communication Networks*. New York, NY: Oxford Press.

Noam, E. M. (2001). *Interconnecting the network of networks*. Cambridge, MA: MIT Press.

O'Hara-Devereaux, M., & Johansen, R. (1994). *Globalwork*. San Francisco, CA: Jossey-Bass.

Parsons, P. (2003). The evolution of the cable-satellite distribution system. *Journal of Broadcasting & Electronic Media*, *47*(1), 1-17.

Parsons, P., & Frieden, R. (1998). *The cable and satellite television industries*. Needham Heights, MA: Allyn & Bacon.

Pilotta, J., Widman, T., & Jasko, S. (1988). Meaning and action in the organizational setting: An interpretive approach. *Communication Yearbook*, *11*, 310-334.

Poole, I. S. (1990). *Technologies without boundaries*. Cambridge, MA: Harvard U. Press.

Porter, M. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York, NY: Free Press.

Stoddard, D., & Jarvenpaa, S. (1995). Business process redesign: Tactics for Managing Radical Change. *Journal of Management Information Systems*, *12*(1),81-107.

Tarn, J. M., Razi, M., Yen, D., & Xu, Z. (2002), Linking ERP and SCM systems. *International Journal of Manufacturing Technology & Management*, *4*(5), 420-439.

Thompson, A., & Strickland, A. (2008). Dell computer corporation, Available at: http://www.mhhe.com/business/management/thompson/11e/case/dell.html. Retrieved: 20 January 2008.

Zheng, S., Yen, D., & Tarn, J.M. (2000). The new spectrum of the cross-enterprise solution: The integration of supply chain management and enterprise resource planning systems. *The Journal of Computer Information Systems*, *41*(2), 84-93.

KEY TERMS

Business Process Innovation: A process is a structured, measured set of activities designed to produce a specified output for a particular customer or market. It implies a strong emphasis on how work is done within an organization. Business process innovation involves creating systems and methods for improving organizational performance.

Cable Networking: HBO's 1975 commitment to advance satellite distribution of programming to cable operating systems ushered in a whole new era specialized television programs, including WTBS, 1976; ESPN, 1979; CNN, 1980; and MTV, 1981. Thus, was born the principle of cable networking; that is, television programming designed exclusively for cable operating systems and later direct broadcast satellite systems.

Enterprise Resource Planning: Enterprise resource planning (ERP) attempts to integrate all departments and functions across an entire company onto a single computer system using a common data base and a shared set of reporting tools. The goal of an ERP system is to replace stand alone programs such as accounting, manufacturing, human resources, warehousing and transportation and replace them with a single unified software program.

Intelligent Networking: Intelligent networking represents the combination of software, technology and electronic pathways that makes business process innovation possible for both large and small organizations alike. A central tenet is that the intelligent network is not one network, but a series of networks designed to enhance worldwide communication for business and residential users.

Just-in-Time Manufacturing: Just-in-time manufacturing means quick turnaround and allows a company to meet an order in the least amount of time. Just-in-time manufacturing (and delivery) relies on the use of supply chain management and ERP systems for the purpose of tracking customer orders.

Reengineering: Reengineering involves a major restructuring (or overhaul) of an organization's key operations. Another term for reengineering is *business process redesign*. There are certain features that are typical of a reengineered process. This includes: 1) Creating cross-functional teams, 2) Streamlining the business process, 3) Designing multiple versions of a business process and 4) Sharing information and resources.

Supply Chain Management: Supply chain management (SCM) is a complex business model that takes into consideration the entire set of linking steps necessary to produce and deliver a product to the end consumer. A supply chain consists of the intermediary steps from the point of product inception (inclusive of the purchase order) to the delivery of the finished good(s) to the consumer. A supply chain is connected by transportation and storage activities and coordinated through planning and networked information activities.

Chapter XXX
Can M-Commerce Benefit from Pervasive Computing?

Stan Kurkovsky
Central Connecticut State University, USA

ABSTRACT

Mobile commerce is a special area of electronic commerce that utilizes mobile wireless devices to conduct commercial transactions. Unique features of these devices include their mobility, personalization, and location-awareness. These features play a very significant role in enabling a novel class of emerging applications that take advantage of recent advances in pervasive computing. The paradigm of pervasive computing environment was first introduced by Mark Weiser in the early 1990s who postulated the anytime anyplace availability of computing and information services that are enabled by miniature devices and sensors seamlessly and unobtrusively embedded in the surrounding environment. Although this vision has not been realized in its literal sense, today mobile and smart phones are commonly viewed as an enabler of a human interface to the surrounding computing and information environment. As an application area where many principles of pervasive computing have been successfully implemented, mobile commerce has reaped significant benefits from the recent scientific and technological advances. This chapter discusses a number of pervasive computing principles and illustrates how they have been implemented in mobile commerce applications. The chapter also presents some new trends in developing context aware m-commerce services that tap into the power of Web 2.0 services and digital communities.

INTRODUCTION

Electronic commerce, or e-commerce, is the process of buying, selling or promoting goods, information, or services via computer networks. Mobile commerce, or m-commerce, is a specialized area of e-commerce, in which mobile devices are used in the process of conducting of a commercial transaction. Unique features of wireless mobile devices have a profound impact on the nature and features of applications designed to run on these devices. Such features include the device ubiquity (mobile phones are affordable and portable), personalization (a typical mobile phone belongs to and can be associated with a single person), and location awareness (a wireless connection can be used to estimate, and a GPS sensor can precisely determine the physical location of the device) (Kannan et al., 2001). Many current e-commerce applications have been adapted for wireless mobile platforms. There is also an emerging class of m-commerce applications, which is enabled by the unique platform features, such as location-awareness and mobility inherent to such mobile devices as smart phones. Such a trend also exists in many other application areas where mobile devices are used: as existing applications migrate to mobile platforms, the features unique to device mobility and wireless connectivity create a unique category of emerging applications aiming to achieve the anytime, anywhere paradigm of pervasive computing. In particular, new m-commerce applications emerge due to three unique characteristics of mobile devices: they provide a communication channel for remote merchants and providers of services through their anyplace anytime connectivity; they provide unique capabilities to personalize the experience of their users by gathering information about the local context; and they allow service providers to affect the user environments via local communication links and actuators (Gershman, 2002). Many researchers believe that m-commerce will continue to serve as a major driving force behind the development of mobile Internet (Maamar 2003, Tamminen et al., 2004).

BACKGROUND

The paradigm of pervasive computing indicates that as technology advances, computing devices will become smaller but more powerful, which would allow these devices to be ubiquitously and invisibly integrated into our everyday surroundings providing an anyplace anytime access to a computing environment. In the early 1990s, Mark Weiser (1991, 1993) described an early prototype of such an environment comprised of three classes of devices: tabs, pads and boards, each of which were designed after the corresponding office instruments. Tabs, similar to Post-It notes, were small location-aware devices with a pressure sensitive screen for writing short notes. Much like today's tablet PCs, pads were wireless pen-based notebooks. Boards were large wall-sized interactive surfaces, functionally similar to office whiteboards. Development of these devices did not progress beyond research prototypes at the Xerox PARC labs. However, this project generated much interest in research and industry due to the impact this paradigm could make on the way how humans interact with technology.

Pervasive computing is an emerging research field full of many open problems. One of the most challenging questions is how to ensure that a computing system is invisibly embedded in the surrounding environment and how to minimize the impact of its possible intrusiveness on perception of the users. In general, there are two complementing solutions to this problem: by using miniature devices and embedding the logic into wearable and mobile devices, and the environment; and by incorporating into the system a degree of intelligence capable of anticipating the user's actions within the current environmental context. Mark Weiser described the paradigm of pervasive computing as a "world where each person is continually interacting with hundreds of nearby wirelessly interconnected computers" (Weiser, 1993) that "weave themselves into the fabric of everyday life until they are indistinguishable from it" (Weiser, 1991). Consequently, a pervasive system designed in this fashion will "fade into the background" and its users will be able to focus on everyday tasks at hand rather than on the idiosyncrasies of interacting with the technology. Much progress has been made in the area of pervasive computing that significantly advanced it from tabs, pads and boards, however, the fundamental theoretical and technological principles supporting pervasive computing remain the same (Kurkovsky, 2007). Of all application areas where the concepts of pervasive computing took strong roots, mobile commerce is one of the domains that benefitted the most from these advances.

The very definition of m-commerce indicates that it is a process conducted by means and with a direct involvement of mobile devices. However, pervasive computing is uniquely positioned to lend its principles to m-commerce and provide a much wider array of features and capabilities furthering the objectives of m-commerce, that is, to increase the number of buying and selling opportunities. Mobile devices do not exist in a vacuum. To begin with, they function in the environment of ubiquitous network connectivity that allows them to fulfill their primary purpose – connect their user to the communications network and the plethora of information services available online. However, there is a growing trend to build some kind of communications capability to many other devices widely available in our surrounding environment. From digital picture frames to kitchen appliances, from cars to retail kiosks, we see an increasing availability of smart appliances and a growing penetration of communication infrastructure into everyday objects. While each of these devices and appliances may have a highly specialized purpose, their connectivity features may make them capable of interfacing with universal and always generic mobile devices. For example, the user accessing a TV schedule on his smart phone should be able to program his TiVo to record the chosen program. The user watching a video clip on a mobile phone should be able to seamlessly switch the video feed to a large screen TV and vice versa. The user using his mobile phone to locate the nearest vending machine should be able to use his phone's capabilities for a payment. As a device that is carried by a person all the time, the mobile phone is uniquely positioned to become the user's interface to a wide variety of pervasive information services surrounding the user. In Weiser's paradigm of pervasive computing, we will be surrounded by ubiquitous information services seamlessly built into the environment. This vision may be far from being implemented in its literal sense, but with help of mobile and smart phones, this vision is already becoming a reality. We can already witness such a transformation in many application areas where pervasive computing has made inroads into our everyday lives. This trend is probably the most visible in the area of m-commerce because developers, wireless carriers, retailers, and information service providers are interested in advancing m-commerce applications to their customers. Given the current demographics of mobile users, early adopters, young people, and technologically-savvy customers are among those who have the highest influence on the adoption trends in mobile applications (Bigné et al., 2007, Pedersen, 2005, Wu & Wang 2005). It has been shown that the very same slice of mobile users has been the most eager to transform a significant portion of their online activities from desktop computers to handheld devices.

PERVASIVE COMPUTING IN MOBILE COMMERCE

Anytime and anyplace availability of pervasive services is provided by modern mobile devices due to their increasingly expanding coverage areas and a growing availability of high-speed data services. Earliest attempt to bring data connectivity to mobile devices were implemented by NTT DoCoMo's i-mode framework and the General Packet Radio Service (GPRS), both of which brought revolutionary advances in the availability of a wide range of online services to the users of mobile devices. Unfortunately, these early attempts suffered from significantly low speeds (comparable to those of dialup modems), which hindered the rollout of data-intensive applications such as web browsing and over-the-air audio and video downloads. Today, however, third generation mobile services (3G) are becoming increasingly available and have achieved a level coverage reaching about 60 to 80 percent of subscribers in many European counties (GSM Europe, 2008). It is worth noting that the rollout of 3G networks and services in the US is significantly slower. With anyplace anytime connectivity provided by their mobile devices, users should be able to tap into m-commerce services that are available to them at any place and at any time. Many of the m-commerce merchants and service providers simply adjusted their e-commerce business models to the mobile market. For example, eBay, the world leading online auction provider, was among the first e-commerce web sites to create a presence on the mobile market (Boulton, 2000). Today, there are many m-commerce services that are enabled by certain unique features of the mobile medium. For example, Verizon Wireless offers a Song ID service on its V CAST handsets, which allows mobile users to identify a song by recording a fragment they might hear some place. While the identification service is free on the compatible handsets, it creates additional sources of revenue due to the increases in the user data traffic and additional selling opportunities when the users are offered to purchase matching ringtones or music tracks.

Mobile commerce, as a natural evolution of electronic commerce, uses mobile devices and their network connectivity to buy, sell, and promote goods, information, and services. Today, there is shift in focus from mere supporting and enabling transactions to the exploiting of enhanced information services available on mobile devices. M-commerce broadens the conventional scope and nature of electronic commerce by incorporating unique features that mobile devices provide to their users – mobility of participation and portability of technology (Elliot & Phillips, 2004). Mobile and smart phones present a unique tri-fold combination of capabilities that make them uniquely positioned to advance and create new and often unexpected m-commerce applications. These capabilities are: a communications channel, a context sensor, and a pervasive interface enabler. Providers of m-commerce services have already discovered that in order to be successful in a highly competitive area, they need to follow three basic principles outlined in Mark Weiser's vision of pervasive computing: to be connected to their customers – anyplace and anytime, to be aware of the customers' context – their activities, surroundings, location, etc., and to be proactive – making their services tailored to the needs and context of each individual customer.

In general, mobile environment can be characterized by its spatiality, temporality and contextuality (Kakihara & Sørensen, 2002). The first two attributes are parallel to the anyplace and anytime availability of network connectivity and information services described by Weiser. Contextuality refers to the way that the owners use their mobile devices while being surrounded by other people, devices, and systems. Context-aware applications take advantage of this property; they use information about the current context to support the decision-making process as a part of improving the quality of the user's experience. A typical context-aware application is able to sense and collect information about different aspects of the user's surroundings, the information about the present and past behavior of the user, and adjust its functionality to better respond to the needs of the user in the current situation. Context-awareness is a growing research area and there is no commonly accepted and universal definition of context; many authors tend to emphasize certain aspects of context over others. However, most research sources agree that context refers to the information about "who you are, where you are, who you are with, and what resources are nearby" (Schilit et al., 1994, Koukia et al., 2006).

In mobile commerce, context information is typically used to reduce the information overload and improve content management by adapting the service, information, and product offerings to the needs of each user in their current situation.

Information overload becomes a critical factor in the framework of the limited resources of mobile devices. Media-rich pages of typical e-commerce portals need to be adapted to small screens of mobile devices and their limited bandwidth. Information on the screen of a mobile device needs to be organized so that it is uncluttered and clearly conveys the most important aspects of the content. Unlike desktop users, mobile users usually have a very limited attention span and any m-commerce service provider will need to compete for the attention of a mobile user with other people, physical advertisements, and other ambient features of the environment (Tarasewich, 2003).

Current trends in m-commerce go beyond simply using mobile devices as mere facilitators of commercial transactions. Instead, through personalization, mobile devices can be used to enable and enhance the likelihood of making a connection of the mobile user wishing to make a purchase and a merchant offering products, services, or information for sale. Context-awareness provides an excellent foundation for providing personalized experiences in the mobile commerce environment. Personalization aims to provide the users with what they need or want without explicitly having to ask or search; context-awareness provides a mechanism to implicitly infer the current needs of the user (Sackmann, 2007, Ala-Siuru & Rantakokko, 2006, Kawsar et al., 2006, Georgiadis et al., 2005).

The focus on mobility in m-commerce applications naturally leads to new avenues and opportunities for connecting merchants and providers of services with their potential customers carrying mobile devices. In a traditional commerce environment, a customer has to be physically present at a merchant's location to make a purchase; in e-commerce applications, a transaction can occur only when the customer is at a computer accessing the web site of the merchant. Mobile commerce breaks the barriers imposed by the physical location of the merchant or the tethered network connection of the customer's computer. M-commerce naturally lends itself to create more serendipitous opportunities for commercial transactions. With anytime availability of online storefronts, mobile customers are able to access them anywhere using their mobile devices. Today we are beginning to witness an emergence of pervasive m-commerce services that fit the description of Weiser's paradigm. Currently, customers can access most modern m-commerce services anywhere and anytime. However, there is an emergence of context-aware and proactive services that are able to sense the user's current surroundings, observe the user's activity, and are capable of making an unobtrusive suggestion, whether it is a purchasing opportunity, an interesting place for sightseeing, or a friend located nearby (Ziv & Mulloth, 2006, Kurkovsky & Harihar, 2006, Nickerson, 2005).

One might say that mobile devices available today are largely "deaf, dumb, and blind" (Gershman, 2002) in the sense that providers of m-commerce services depend on the customers to initiate the transactions by visiting a merchant's web site, uploading a photo to be printed, or entering any other information that would lead to commencing of a commercial transaction. However, more and more devices come equipped with one or more sensors that enable these devices to "hear and see" what is going on around them. The vast majority of current mobile devices are equipped with cameras capable of taking still pictures and videos. Until recently, these cameras have not been used for anything other than for their most immediate purpose. The possibilities, however, are nearly endless. Camera-equipped mobile devices can be used for estimating the changes in the user's position using a continuous video feed (Miyaoku et al., 2004); as barcode scanners (Rohs & Gfeller, 2004); for providing real-time updates to photo albums, photo blogs, and travel diaries (Ashbrook et al., 2006); for real-time scanning of small documents (ScanR, 2008). Ballagas et al. (2006) published a detailed survey of many other applications of cameras in mobile devices.

Built-in GPS sensors provide mobile devices with simple and precise geo-positioning capabilities needed to provide the location and position context features of m-commerce applications. For example, a vehicle equipped with a navigation system capable of sensing the current level of gasoline could inform the driver when the vehicle approaches a gas station selling the preferred brand of fuel. Many practical m-commerce applications enabled by location context have been described in the literature. For example, SMMART (System for Mobile Marketing: Adaptive, PeRsonalized and Targeted) delivers narrowly targeted promotional information to the users of wireless mobile devices. The system uses location context to deliver promotions or coupons to the users when they are in a close proximity or inside a retail site distributing promotions. SMMART also uses other types of context information; the system adapts to the needs of its user by unobtrusively monitoring their shopping habits and learning the user's personal preferences (Kurkovsky & Harihar, 2006).

From the perspective of many m-commerce applications, mobile phones are often viewed as their owners' identity in the sense that each mobile phone is typically owned by a single person, that person carries the phone with them most of the time, and that same phone is used by its owner to conduct all m-commerce transactions. Mobile phones can also be used to store bank and credit card account numbers, as well as login information for remote authentication with m-commerce sites. NTT DoCoMo introduced its FeliCa authentication technology that is used for close proximity authentication applications; one of the earliest applications of this kind was using FeliCa-enabled mobile phones as passes in Japanese mass transit systems. As mobile devices become increasingly associated with their owners who store their personal and sensitive information on the mobile phones, the issue of security and privacy becomes ever more important. Biometric sensors such as retina or fingerprint scanners can be easily used for identification and authentication of the user and preventing any unauthorized access to the mobile device. As described above, cameras built into mobile phones are being used in an increasingly rich array of new applications. It is just a matter of time until we see new applications that use mobile phone cameras as retina scanners to identify the rightful owner of the mobile device.

One of the most appealing features of context-aware systems is the ability of sense surrounding objects. RFID technology provides a relatively simple solution for implementing this aspect of context awareness (Want, 2006). Initially, lack of standards led to a certain fragmentation of the market and resulted in a limited adoption of the RFID technology. However, many early applications became sustainable, such as the EZ-Pass highway toll system. Recently adopted standards and drastic reduction of RFID tags and sensors costs will lead to a wider adoption of this technology in many areas, including m-commerce applications (Nath et al., 2006). NTT DoCoMo's FeliCa system mentioned above uses near-field RFID tags embedded in mobile phones. Currently, a number of brick-and-mortar stores have already implemented RFID-based inventory control and payment systems. One of the most straightforward implementations involves attaching RFID tags to all items sold in a store and using static or mobile RFID scanners for checkout. This scenario could be further enhanced if customers could use their mobile devices as electronic wallets, which can be easily enabled by building an RFID tag into the device.

Due to the advances in manufacturing technologies, mobile devices become ever more capable of performing complex tasks and possess an ever increasing array of features. However, due to their small size and practical considerations of usability, it is doubtful that we will soon see a wide proliferation of mobile devices embedded with all of the above-mentioned and other sensors. On the other hand, most current mobile devices come equipped with Bluetooth connectivity, which, essentially, allows them to become communication hubs for a virtually unlimited number of other specialized devices and sensors. For example, one of the most popular applications of this nature available today is using a Bluetooth-enabled vehicle navigation system in conjunction with a mobile phone, which is used to receive real time traffic updates that are then displayed on the screen of the GPS unit and used in routing calculations.

THE HUMAN FACTOR

There are a wide variety of factors that define the success of a mobile technology; instant connectivity, convenience, and personalization are among the most important (Mahatanankoon & Garcia, 2007). Successful mobile applications are typically perceived by their users as being accessible anytime and anywhere, highly adaptable to the user's needs, and not revealing any private information about the user. However, many potential consumers of mobile technology are not convinced that all m-commerce applications would provide them with such an experience. Consumers are often wary about engaging in m-commerce transactions because many m-commerce features and services are not identical to those of e- commerce. Presently, m-commerce is not positioned to replace e-commerce; instead it is best suited to supplement it. A number of socio-technological factors have a significant effect on the customer perception and the rate of adoption of mobile technology and m-commerce. These factors form tightly integrated socio-psychological barriers that include user unawareness, inefficiency of devices, security and rivacy concerns, and others.

Consumers are not always aware of the capabilities of their mobile devices to conduct m-commerce transactions, availability of m-commerce services, and/or the pricing scheme of their wireless carrier. With a wide range of third-party e-commerce vendors, it is frequently left to the initiative of the users to discover the possibilities for m-commerce transactions. In many scenarios of using mobile devices for commercial transactions, self-efficacy of the consumer plays a very important role in exploring new functionalities.

Limited functionality of mobile devices may further hinder the willingness of the users to actively use them as m-commerce medium. Research indicates that every additional navigational input reduces the chances of completing of a transaction by half (Clarke, 2001). Limited screen size, battery power, and processor speed require device manufacturers to be acutely aware of the multitasking behavioral patterns of the device users (Lee & Benbasat, 2003). Personalization and adaptability of applications can typically compensate for the ergonomic drawbacks and limitations of the user interface, but it may adversely impact other aspects of the users' experience, such as privacy and security. There is a clear need for mobile applications to be able to accommodate the limited attention span of mobile users and the patterns of their multitasking behavior.

Personalized features of context-aware applications are typically more user-friendly in that they can adapt themselves to the needs of each individual user. However, most context-aware and customizable systems require that the users provide some personally-identifiable information and/or allow the system to track their behavior and actions. This inevitably leads to the issues of privacy and security, which are a major concern for mobile users. Obviously, concerns for individual privacy have a negative impact to the adoption of m-commerce services. Consumers rightfully fear that they can be tracked and profiled, and their transaction history and navigation patterns can be analyzed and misused (Pitkow et al., 2002). As a result, lack of trust to m-commerce providers may also lead to consumers avoiding the use of personalization/customization features in m-commerce applications.

Typical characteristics of traditional brick-and-mortar and e-commerce can have a significant impact on the way how consumers perceive m-commerce. Prior experience with e-commerce services may also impact the tendencies of the consumers to adjust their behavior to the nature of mobile devices (Orlikowski & Gash, 1994). Trustworthiness of m-commerce services still remains questionable. The very idea of not being able to deal with a live person face to face, or not being able to physically touch the merchandise may not sound attractive to consumers (Mahatanankoon & Garcia, 2007). A large population of mobile users do not welcome the idea of entering personal and financial information into their mobile devices to complete an m-commerce transaction because they feel that wireless networks may be insecure or fear becoming a victim of identity theft if their mobile device is lost or stolen. E-commerce services often welcome users to participate in virtual communities where users can interact with other customers and share their opinions about the products; most current m-commerce services lack anything even remotely similar.

These socio-technological factors suggest that the user experience in the mobile environment is drastically different from that of the traditional or electronic commerce. Customers of an e- commerce site may make a decision to buy a product from that vendor simply by researching its reliability and customer reviews. Presently, there are not many convenient possibilities for m-commerce customers to use a similar functionality not only because customer review or virtual community services may be limited for m-commerce vendors, but simply because it may be very inconvenient for the user to multitask and switch between different screens required to complete such a research task. The m-commerce industry must help its customers to overcome these barriers before any m-commerce services can become significantly profitable.

FUTURE TRENDS

A possible progression along the evolutionary chain of brick-and-mortar commerce to electronic commerce and to mobile commerce, is the integration of the elements of social context (Maamar, 2003). What customers often lack in e-commerce and m-commerce applications, is the sense of presence where they can physically touch and experience the products before they purchase it. Customers also often lack the sense of community they may feel when shopping in a physical store where they can interact with other customers and perhaps ask for their advice or opinion. Designing m-commerce and e-commerce applications that simulate the feeling of being in a real world marketplace can dramatically increase the user confidence and willingness to purchase goods or services. The trend to incorporating such features is evident, especially, in e-commerce applications. For example, amazon.com has long used customer reviews as an avenue to improve the user experience and give shoppers a chance to get an unbiased advice from others. Recommendations based on purchase history of an individual shopper matched against a range of other shoppers who purchased similar products provides a significant motivation for customers to explore new products while having a high likelihood of being interested in such offerings. These trends also migrate to the mobile commerce arena, where they can be further enhanced with the capabilities of mobile (e.g. Dodgeball described by Ziv & Mulloth, 2006) and traditional (e.g. Facebook's Beacon) social networks. Video sharing portals, such as YouTube, begin provide offerings especially tailored to mobile devices; many merchants are taking advantages of these capabilities by allowing their customers to view video product demos on their mobile devices.

Evolution in the area of dashtop computing should also be noted as a very promising area that is already seeing a limited introduction of m-commerce services. In general, dashtop computing describes any kind of computing applications embedded in vehicles, primarily associated with navigation systems and telematics (Bisdikian et al., 2002). Dashtop platforms add a significant amount of convenience and safety features to the drivers. Modern dashtop and embedded telematics systems typically include a GPS navigation system, Bluetooth connectivity for interfacing with other devices (most importantly, with cell phones), and are usually interfaced with other computerized vehicle systems. As a result, such systems are capable of performing most operations hands free, which is crucial to maintaining safe driving conditions. Typical m-commerce applications include live traffic updates used for real time routing, vehicle service reminders with suggestions on related services and products, vendor proximity notifications, and other value-added services.

Current market research indicates that the main source of growth in e-commerce sites comes from the increased spending of existing customers exceeding that of the new customers (Lin, 2008). As a consequence, successful online retailers must be able to retain their existing customers, motivate them to spend more, and potentially take existing customers away from the competitors. One of the ways to achieve these goals for traditional e-commerce merchants is to embrace the mobile market and harness emerging online social networking applications ant technologies, which are commonly termed Web 2.0. Digital communities created a new business model that provides free content creation and sharing facilitated by such Web 2.0 sites as MySpace, YouTube, Flickr, and FaceBook. Content created by individuals can quickly become popular as it is accessed by millions of visitors. By generating enormous traffic, such Web 2.0 sites have created significant opportunities for advertising and branching out to the mobile market with podcast, advertising, and targeted marketing campaigns.

Online advertising market is an extremely lucrative area. Several years ago, many analysts predicted that online advertisement will make a very quick progress in conquering the mobile market (Leske, 2008). However, mobile advertising has been very slow to take roots in mobile medium primarily due to the unique features imposed by mobile devices and the usability characteristics dictated by the limited attention span of mobile users and limited resources of mobile devices. To be successful in the mobile market, advertisers must ensure that they reach the right people with the right advertisements at the right time (Ranganathan & Campbell, 2002). Furthermore, advertisers need to ensure that their ads are delivered in the best possible format depending on the bandwidth and type of connectivity of each individual mobile device.

Speech interface remains the most natural way that humans use to communicate. Furthermore, all mobile phones support voice communications, no matter how complex or simple they are. There many domain-specific applications facilitated by such technologies as VoiceXML and SALT (Kurkovsky et al., 2008); most of these applications were originally used for online transactions and information retrieval using landline telephony. However, the very same applications are immediately accessible via mobile phones. Furthermore, user mobility and independence of the tethered line of communication may create a number of unique capabilities and features specific to the domain of m-commerce. For example, voice interface has been successfully implemented in a wide range of location-based systems including directions and navigation (Bisdikian et al., 2002), sightseeing (Nickerson, 2005), and recommender systems (Sharp & Kurkovsky, 2007).

Despite its tremendous promise, m-commerce is facing significant acceptance barriers, which are caused primarily by the user interface restrictions of mobile devices (Fan et al., 2005). Generally, mobile users find it inconvenient to scroll through large amounts of information using a small screen and often inconvenient keypads. Current research shows that each input action that the user is required to make while using a mobile device decreases by half the chances of making an m-commerce transaction (Clarke, 2001). Speech interface, however, have the capability to alleviate many of the limitations of the visual interface on mobile devices.

CONCLUSION

From an emerging research area a decade ago, pervasive computing has matured enough to make contributions to many other areas of computing and information technology. Today we can witness how the advances made in pervasive computing research are reaching across the boundaries of subject domains. It is likely that mobile commerce has reaped the most benefits from these new advancements. Context awareness features are viewed by the providers of mobile commerce services as a significant factor in increasing the amount of their sales, keeping a loyal base of customers, as well as attracting new ones. Customers also appreciate many features that context awareness can provide: typically, context-aware m-commerce services are able to provide a significantly higher degree of personalization and serendipity, making highly relevant suggestions and recommendations to the customers at the right place at the right time. The paradigm of pervasive computing envisioned by Mark Weiser over 25 years ago encompasses the anyplace anytime access to information and computing services. Today, these capabilities went far beyond being an abstract laboratory experiment and became a tangible reality through their implementation in mobile commerce and many other application areas.

REFERENCES

Ala-Siuru, P., & Rantakokko, T. (2006). Understanding and recognizing usage situations using context data available in mobile phones. *2nd International Workshop on Personalized Context Modeling and Management for UbiComp Applications*. Berlin: Springer-Verlag.

Ashbrook, D., Lyons, K., & Clawson, J. (2006). Capturing Experiences Anytime, Anywhere. *Pervasive Computing*, *5*(2), 8-9.

Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The Smart Phone: A Ubiquitous Input Device. *Pervasive Computing*, *5*(1), 70-77.

Bigné, E., Ruiz, C., & Sanz, S. (2007). Key drivers of mobile commerce adoption an exploratory study of Spanish mobile users. *Journal of Theoretical and Applied Electronic Commerce Research*, *2*(2), 48-60.

Bisdikian, C., Boamah, I., Castro, P., Misra, A., Rubas, J., Villoutreix, N., Yeh, D., Rasin, V., Huang, H., & Simonds, C. (2002). Intelligent Pervasive Middleware for Context-Based and Localized Telematics Services. *2nd International workshop on mobile commerce* (pp. 15-24). New York, NY: ACM Press.

Boulton, C. (2000). eBay joins mobile fray. *InternetNews.Com*, *May 22, 2000*. Retrieved on February 15, 2008, from http://www.internetnews.com/ec-news/article.php/376621

Clarke III, I. (2001). Emerging value propositions for m-commerce. *Journal of Business Strategies*, *18*(2), 133-148.

Elliot, G., & Phillips, N. (2004). *Mobile Commerce and Wireless Computing Systems*. Upper Saddle River, NJ: Addison Wesley.

Fan, Y., Saliba, A., Kendall, E. A., & Newmarch, J. (2005). Speech Interface: An Enhancer to the Acceptance of M-commerce Applications. *International Conference on Mobile Business*. Washington, DC: IEEE Computer Society.

Georgiadis, C., Mavridis, I., & Manitsaris, A. (2005). Context-Based Humanized and Authorized Personalization in Mobile Commerce Applications. *International Journal of Computing & Information Sciences*, *3*(2), 1-9.

Gershman, A. (2002). Ubiquitous commerce – always on, always aware, always pro-active. *2002 Symposium on Applications and the Internet*. Washington, DC: IEEE Computer Society.

GSM Europe. (2008). 3GSM Licences: Overview of the Situation in the different EU Member States. *GSM World*, Retrieved February 15, 2008, from http://www.gsmworld.com/gsmeurope/faq/3g.shtml

Kakihara, M., & Sørensen, C. (2002). Mobility. *35th Annual Hawaii International Conference on System Sciences*. Washington, DC: IEEE Computer Society.

Kannan, P., Chang, A., & Whinston, A. (2001). Wireless Commerce: Marketing Issues and Possibilities. *34th Hawaii International Conference on System Sciences*. Washington, DC: IEEE Computer Society.

Kawsar, F., Fujinami, K., Pirttikangas, S., & Nakajima, T. (2006). Personalization and Context Aware Services: A Middleware Perspective. *2nd International Workshop on Personalized Context Modeling and Management for UbiComp Applications*. Berlin: Springer-Verlag.

Koukia, S., Rigou, M., & Sirmakessis, S. (2006). The Role of Context in m-Commerce and the Personalization Dimension. *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. Washington, DC: IEEE Computer Society.

Kurkovsky, S., & Harihar, K. (2006). Using Ubiquitous Computing in Interactive Mobile Marketing. *Personal and Ubiquitous Computing, 10*(4). Berlin: Springer-Verlag.

Kurkovsky, S. (2007). Pervasive Computing: Past, Present and Future. *5th International Conference on Information and Communication Technology*. Washington, DC: IEEE Computer Society.

Kurkovsky, S., Strimple, D., Nuzzi, E., & Verdecchia, K. (2008). Mobile Voice Access in Social Networking Systems. *5th International Conference on Information Technology: New Generations*. Washington, DC: IEEE Computer Society.

Lee, Y. E., & Benbasat, I. (2003). Interface design for mobile commerce. *Communications of the ACM*, *46*(12), 49-52.

Leske, N. (2008). Mobile ads take baby steps to reach consumers. *Reuters Technology news*, Retrieved on February 16, 2008 from http://www.reuters.com/article/technologyNews/idUSL1445372720080214

Lin, K.-J. (2008). E-Commerce Technology: Back to a Prominent Future. *Internet Computing, 12*(1), 60-65. Washington, DC: IEEE Computer Society.

Maamar, Z. (2003). Commerce, E-Commerce, and M-Commerce: What Comes Next? *Communications of the ACM, 46*(12), 251-257. New York, NY: ACM Press.

Mahatanankoon, P., & Garcia, J. (2007). Mobile Commerce Adoption Barriers. *Encyclopedia of Mobile Computing and Commerce*. IGI Global.

Miyaoku, K., Higashino, S., & Tonomura, Y. (2004). C-Blink: A Hue-Difference-Based Light Signal Marker for Large Screen Interaction via any Mobile Terminal. *17th Annual ACM Symposium on User Interface Software and Technology*. New York, NY: ACM Press.

Nath, B., Reynolds, F., & Want, R. (2006). RFID Technology and Applications. *Pervasive Computing, 5*(1), 22-24. Washington, DC: IEEE Computer Society.

Nickerson, M. (2005). History Calls: Delivering Automated Audio Tours to Visitors' Cell Phones. *2005 IEEE Conference on Information Technology: Coding and Computing*. Washington, DC: IEEE Computer Society.

Orlikowski, W. J., & Gash, D. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems*, *12*(2), 174-207.

Pedersen, P. E. (2005). Adoption of Mobile Internet Services: An Exploratory Study of Mobile Commerce Early Adopters. *Journal of Organizational Computing and Electronic Commerce*, *15*(3), 203-222.

Pitkow, J., Schutze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., et al. (2002). Personalized search. *Communications of the ACM*, *45*(9), 50-55.

Ranganathan, A., & Campbell, R. H. (2002). Advertising in a Pervasive Computing Environment. *2nd international workshop on mobile commerce*. New York, NY: ACM Press.

Rohs, M., & Gfeller, B. (2004). Using camera-equipped mobile phones for interacting with realworld objects. In A. Ferscha, H. Hoertner, G. Kotsis (Eds.), *Advances in Pervasive Computing* (pp. 265-271). Vienna, Austria: Austrian Computer Society.

Sackmann, S. (2007). Personalization and Privacy in Ubiquitous Computing – Resolving the Conflict by Legally Binding Commitments. *9th IEEE International Conference on E-Commerce Technology*. Washington, DC: IEEE Computer Society.

ScanR (2008). Retrieved on February 15 from http://www.scanr.com

Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications, *1st International Workshop on Mobile Computing Systems and Applications*. Washington, DC: IEEE Computer Society.

Sharp, A., & Kurkovsky, S. (2007). VOICE: a Framework for Speech-Based Mobile Systems. *2nd IEEE International Symposium on Pervasive Computing and Ad Hoc Communications*. Washington, DC: IEEE Computer Society.

Tamminen, S., Oulasvitra, A., Toiskallio, K., & Kankainen, A. (2004). Understanding Mobile Contexts. *Personal and Ubiquitous Computing*, *8*(2), 135-143.

Tarasewich, P. (2003). Designing Mobile Commerce Applications. *Communications of the ACM*, *46*(12), 57-60.

Want, R. (2006). An Introduction to RFID Technology. *Pervasive Computing*, *5*(1), 25-33. Washington, DC: IEEE Computer Society.

Weiser, M. (1991). The Computer of the 21st Century. *Scientific American*, *265*(3), 66-75.

Weiser, M. (1993). Some Computer Science Issues in Ubiquitous Computing. *Communications of the ACM*, *36*(7), 75-84.

Wu, J.-H., & Wang, S.-C. (2005). What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. *Information and Management*, *42*(5), 719-729.

Ziv, N. D., & Mulloth, B. (2006). An Exploration on Mobile Social Networking: Dodgeball as a Case in Point. *International Conference on Mobile Business*. Washington, DC: IEEE Computer Society.

KEY TERMS

Context: Any information characterizing the situation of a person, a place, or an object that is relevant to the interaction between the computing system and its user.

Context Awareness: A property of a computing system that can adapt its behavior to the current context without any explicit user intervention.

Mobile Commerce: A special area of electronic commerce that utilizes mobile wireless devices to conduct commercial transactions.

Mobile Device: A pocket-sized computing device that typically has a small screen for user output and a miniature keyboard or touch screen for user input.

Mobile Phone: A mobile device used for voice communication that uses a cellular network.

Pervasive Computing: A model of human-computer interaction in which information processing is seamlessly integrated into everyday objects and activities allowing the user to concentrate on the specific task rather than on details of interacting with the system.

RFID (Radio-Frequency IDentification): An automatic identification method, relying on storing and remotely retrieving data using RFID tags or transponders.

Smart Phone: A mobile phone that offers advanced capabilities that combine those of a typical mobile phone with PC-like functionalities.

Section IV
Telecommunications Technology Management

Chapter XXXI
Basics of Telecommunications Management

Katalin Tarnay
University of Pannonia, Hungary

ABSTRACT

The purpose of this chapter is to provide a simple understandable approach to the basics of telecommunications management. This chapter is composed of four logically interlaced parts: the background information, the main components of telecommunications management, the future trends, and the conclusion. The background information section reviews the literature. The main part discusses management notions together with a short explanation of basic management systems. These systems include the management functions, the managed objects, the management information base, and the management protocols. The future trends section provides solutions most likely to appear in the future. The conclusion summarizes the fundamental mechanisms of this field.

INTRODUCTION

The worldwide information exchange and the mobility for users as well as for devices resulted in the integration of different telecommunication networks. The first telecommunication networks were wired and regional. For example, in Europe, regionalism meant the domination of local national wired networks in each different country. This structure only changed around the end of the 20th century with the appearance of new technologies that allowed the extension of regional network groups into a huge, global wired and even wireless network. The most important global network is the Internet. The basic idea about the Internet was to create the 'network of networks', where users all around the world can be integrated into one huge network. Therefore the main purpose of the Internet is to allow users to communicate with each other in a fast and reliable way. The next big step forward was the creation of the World Wide Web. This let users to reach various services, like Internet based databases or applications. One of the biggest successes around the Web was the development of the electronic business, or simply e-business.

This evolution requires new telecommunications management methods, which are pointing towards the development of the fundamentals of telecommunications management, first of all, the management protocols and Management Information Bases (MIBs). Advanced protocols and well-structured MIBs are the key components to ensure a fast and secure network operation. At the same time, the greater the network, the more management problems may occur. One serious management problem is the lack of security. For example, it is possible that a user can reach services or information which he or she was not authorized for.

Early management systems were responsible only for fault detection. After the networks became wider and the numbers of users increased, the necessity of other management functions emerged. There are five management functions today: fault, account, configuration, performance, and security management. These functions are associated with the network resources, called the managed objects.

The objective of this chapter is to introduce and explain the basics of telecommunications management focusing on the user's point of view. Therefore, this chapter explains the fundamentals of telecommunications management in an easily understandable way: starting at the definition of telecommunications management, presenting the management functions, the managed objects, the management information bases and the main management protocols. Towards the end of the chapter, a simplified example on telecommunications management functions demonstrates the operation of a distributed network, and then the future trends focus on the expected new techniques and solutions, while the conclusion summarizes the contents of this chapter.

BACKGROUND

Network management takes place between two major types of systems: those in control, called managing systems, and those observed and controlled, called managed systems. Majority of the managing systems are called Network Management Systems (NMS). Managed systems can include hosts, servers, or network components such as routers or intelligent repeaters. To promote interoperability, cooperating systems must adhere to a common framework and a common language, called a protocol. In a managed device, specialized low-impact software modules, called agents, access information about the device and make it available to the NMS. Managed devices maintain values for a number of variables and send reports about those, as required, to the NMS. For example, an agent might report such data as the number of bytes and packets in and out of the device, or the number of broadcast messages sent and received. In the Internet network management framework, each variable is referred to as a managed object, which is anything that an agent can access and report back to the NMS.

All information about the managed objects is contained in the Management Information Base (MIB), which is a database of the managed objects. The managed objects, or variables, can be set or read to provide information on network devices and interfaces. An NMS can control a managed device by sending a message to an agent of a managed device requiring the device to change the value of one or more of its variables.

There are two main methods to describe telecommunications management (Stallings, 2002; Carr & Snyder, 2002). Our explanation follows a top-down approach. We start from the main objective of the management, to show how the resources can be managed, controlled and coordinated to reach high-level and continuous services. The other approach discusses the subject in reverse, from bottom up. It starts with the protocols, either from the Simple Network Management Protocol or the Common Management Information Protocol. Both discussion methods have their advantages: The top-down approach is nearer to application management and to the 'user's world', while the bottom-up approach has more benefits for service providers. Because this paper is mainly addressed to users, therefore the top-down approach was chosen.

MAIN COMPONENTS OF TELECOMMUNICATIONS MANAGEMENT – DEFINITION OF TELECOMMUNICATIONS MANAGEMENT

Telecommunications management means monitoring, controlling and coordinating system resources. These networks elements are responsible for maintaining the reliable communication in the network. System resources are identified as managed objects, also called clients or agents (Buchanan, 1999). Monitoring is to query the values of the main attributes or general operational parameters (e.g. layer identifier, timer identifier and window limit) of managed objects. Controlling is to set these to desired values according to some requirements. Controlling a particular managed object is independent of controlling all the other managed objects in the network. Monitoring and controlling are based on the dialogue between the agent and a special host responsible for the management functions, called the manager (sometimes also referred to as the management station). Because many dialogues can take place simultaneously, the results of these dialogues have to be coordinated.

The simplified view of telecommunications management contains a management station as a host and all its managed objects. The management station has a layered structure. Its application layer is connected to each managed object through telecommunications management protocols. Every necessary telecommunications management protocol is realized as a dialogue. The application layer is also connected to a Management Information Base (MIB) containing all the information about the managed network (Garg, 2002).

In the telecommunications sector, a five-layer architecture, the Telecommunications Management Network (TMN), was created by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) to manage business processes (International Telecommunication Union, 2000). TMN layers are often represented as a pyramid-like structure, the layers in the pyramid from bottom to top are: the physical elements of the network, the management of network elements, the network management, the service management, and on the top of the pyramid, the layer for business process management. The management entities are commonly represented in one of the layers above, and belong to the following management functional areas: fault management, account management, configuration management, performance management, and security management.

TELECOMMUNICATIONS MANAGEMENT FUNCTIONS

Fault Management

Fault management is a set of functions, which enables the detection, the isolation and correction of abnormal operation of the telecommunications management network and its environment. There are two main cases whenever a network may detect an error. In the first case, the value of an attribute of an object is out of limit. In the second case, the object is out of operation. The actual fault management is composed of two main and a third optional steps. In the first step called fault localization, the fault is logged and the exact location of the fault is determined. In the second step, the management station runs diagnostic tests, then isolates the fault, and reconfigures the network excluding the faulty device. A possible third step is the improvement of the fault. The fault management is often connected with an alarm mechanism. The alarm mechanism contains alarm reports with information about alarm type error identifiers and possible reasons. The quality assurance measurements for fault management include component measurements for Reliability, Availability and Survivability (RAS).

Configuration Management

Configuration management discovers the location and the name of the managed objects and stores them in the Management Information Base supplemented with all necessary operational parameters. The functions of configuration management are: the name management, the data collection, the parameters setting and initialization. The configuration management deals with network planning and engineering, software installation, status and control checks, service planning and negotiation, and also provisioning. In other words, adding of new programs, new equipment, modification of existing systems, and the removal of obsolete systems and programs belong to configuration management.

Account Management

Account management (or accounting management) deals with the cost of components, such as resources and the service costs, tariffs to determine the cost of the usage of managed objects and the discovery of any account frauds. The account management also deals with the collection and evaluation of billing information. This helps to minimize the cost of operations by making the most effective use of the systems available. Account management is also responsible for ensuring the appropriate billing for users. The accounting process is a particularly complex task in cross-continental wired and wireless networks.

Performance Management

Performance management controls the managed objects from the point of view of cost efficiency, that is, the general behavior of telecommunications equipments and the effectiveness of the network or network elements. Its role is to gather and analyze statistical information for the purpose of monitoring and correcting the behavior and effectiveness of the network, and to aid planning, provisioning, maintenance and quality measurements. Performance management also logs attribute values and monitors response times to ensure the greatest overall performance. It supports customized network settings and collects reports about the network's performance.

Security Management

Security management deals with the protection of the managed objects by distributing and handling security services. Its functions include security services for communications and security event detection and reporting. The security services for communications provide services for authentication, authorization, access control, data confidentiality, data integrity and non-repudiation that may occur in the course of any communications between systems, between customers and systems and between internal users and systems. The function of security event detection and reporting reports to higher layers of security any activity that may be interpreted as a security violation (e.g. unauthorized user, physical tampering with equipment).

MANAGED OBJECTS AND MANAGEMENT INFORMATION BASES

Managed Objects

Managed objects are abstract representations of network resources, the attributes of network resources related to management. All managed objects are represented in the Management Information Bases (MIBs). The components that must be represented in the MIBs are: The attributes related to system resources, the management operations related to all managed objects (create, delete and action), the management operations related to the value of the attributes (get, set, derive, add and remove) and the responses to the management operations and the reports or event reports, which can be outer events (a connection was lost with a neighbor component) or inner events (a counter has overrun a limit value). All groups of managed objects can be considered as a hierarchical structure.

The highest level in the hierarchy is the system. The system includes the hardware components, the operating system, the network software and its version number, the sub network group name and the exact time of initializing the managed system. Below the system, the interface is presented. The interface contains the number of network interfaces, the size of the datagram or packet which the interface can receive, the bandwidth of the interface, the address of the interface, its operational state and the features of the traffic. On the same level, next to interfaces, we have the address translation, which maps the network address to a hardware device's address. Below this level, the protocols are presented. The most important are the followings: Internet Protocol (IP), Transmission Control Protocol (TCP), Stream Transmission Control Protocol (STCP), User Data Protocol (UDP), Exterior Gateway Protocol (EGP) and Simple Network Management Protocol (SNMP). The managed objects and the Management Information Bases have the same hierarchical structure.

Management Information Base (MIB)

The Management Information Base (MIB) is an integrated set of the abstract representation of all managed objects, which is used in open systems to manage the devices in a communications network (McGinnis & Perkins, 1996; Ray & Abbi, 2004). It comprises of a collection of objects in a virtual database used to manage entities (such as routers and switches) in a network. Within any given MIB, the Structure of Management Information (SMI) defines the format for all objects maintained in that MIB. The SMI is divided into three parts: the module definition, where the information modules are described, the object definition, where the managed objects are described and the notification definition, where the transmissions of management information are described. The SMI defines the objects formats using a particular form of notation called Abstract Syntax Notation One (ASN.1) (International Telecommunication Union, 2002a & 200b).

ASN.1 is a data representation used in every MIB. This definition of the data representation provides rules for handling heterogeneous forms of data representations, like communication of machines with different data formats. It also defines the encoding of abstract values into a byte stream for the transmission through a network. The different data representations between the sending and the receiving machines should be handled. There are two possibilities for this method. Either one of the machines converts the data into the format of the other, or the sending or the receiving machines both convert data into a standard form. The first method will result in many machines with different data formats and many conversions, which is quite disadvantageous. Therefore the second method is the generally accepted data representation format. ASN.1 is a standard notation that allows one to define a wide range of data types and structures. It uses two types of syntaxes, the abstract syntax which is a set of rules that specify data types and structures for information storage, and the transfer syntax which is a set of rules for communicating information between systems. ASN.1 is widely used in networking to define data of higher level protocols.

TELECOMMUNICATIONS MANAGEMENT PROTOCOLS

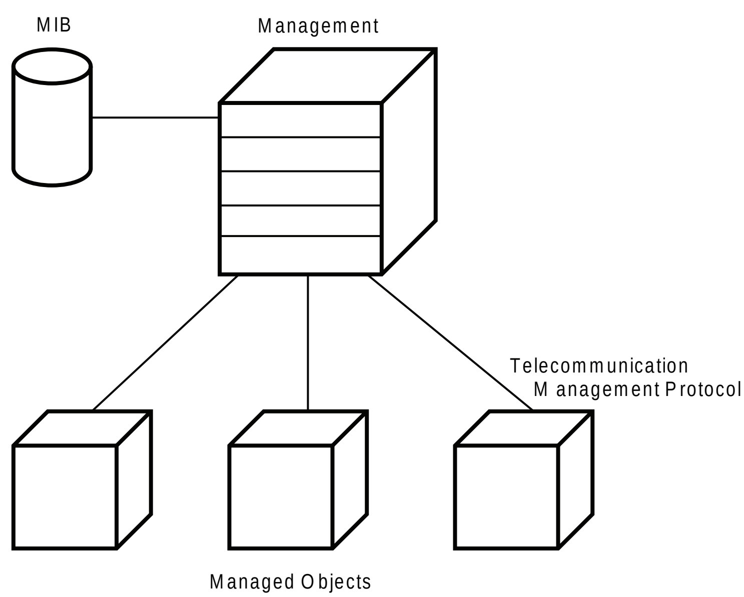
Telecommunications management protocols are responsible for the network management mechanisms and they are classified in three groups: monitoring, controlling and coordination. Monitoring means to observe the whole operation of the system. This process is also responsible for detecting the exact place and type of any errors. Controlling means to set the required values of the attributes, where the attributes are the variables of the managed objects. Coordination means to harmonize the operation of the managed objects.

The telecommunications management protocols represent a logical connection between managed objects and management systems (see Figure 1). In this section we will discuss the first significant management protocol called the Simple Network Management Protocol (SNMP), then the Open System Interconnection management protocols, and finally the main management protocols for the mobile Internet. We will also provide a brief description of authentication, authorization and accounting.

Simple Network Management Protocol (SNMP)

SNMP has three versions: SNMP, SNMPv2 and SNMPv3. SNMP has a modular and extensible architecture and supports remote configuration (Stallings, 1998). Being a simple request-response protocol, it uses the following commands: 'get', 'getNext', 'set' and 'trap'. 'Get' and 'getNext' is used for reading object information, while the 'set' command is used to give a new value to the requested attribute. 'Trap' is for monitoring faulty events. The data representation is ASN.1 based and SMI specific. Connecting wide area networks, there is a great probability to have different SNMP versions, e.g. SNMPv1 and SNMP v2c. In this case, the coexistence between these two different versions is allowed. SNMP is not tied to any particular set of data structures; it operates on a collection of related objects identified in the SNMP MIB. Objects in the MIB are described by ASN.1 naming scheme using a hierarchical naming structure. For efficiency, each name has a numeric equivalent. The SNMP MIB basically has two types: MIB 1 and MIB 2. The SNMP MIB

Figure 1. Main components of telecommunications management



1 was the original TCP/IP standard for protocol suite, while SNMP MIB 2 extends that original version. In the SNMP MIB, the object identifier is used to refer to single MIB object. The object identifier is a sequence of non-negative integers that traverses an object tree. The tree starts with the root, while the branches of the object tree are referred to as subordinates. The objects are identified by writing the path used to get to a specific device. As a result of the original SNMP version, sometimes referred as SNMPv1, had poor security and billing capabilities, further development proved to be necessary. SNMPv2 revises version 1 and includes improvements in the areas of performance, security, confidentiality, and manager-to-manager communications. SNMPv3 furthermore provides three important services: authentication, privacy and access control.

Open System Interconnection Management Protocols

The Open System Interconnection (OSI) model gives a clear and unambiguous guide for cooperation of different communicating systems. The OSI network model describes the communication divided into seven layers. Each layer performs a specific task. These tasks are fulfilled by software and hardware components. The components of different systems can be connected in an open manner. The main motivation is to produce an easy-to-realize solution for the diversity of heterogeneous components.

Different vendors develop different products. The essential part of the OSI model is the Basic Reference Model developed both by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) and the International Standardization Organization (ISO). This abstract model of networking is a seven-layer model. A layer is a collection of related functions which provides services to the above layer and receives services from the layer below. Each layer contains a set of protocols fulfilling many communication tasks. Parts of the OSI model have influenced the Internet protocol development. These Internet protocols are classified into a four-layer abstract model, which is called the TCP/IP model developed by the Internet Engineering Task Force (IETF).

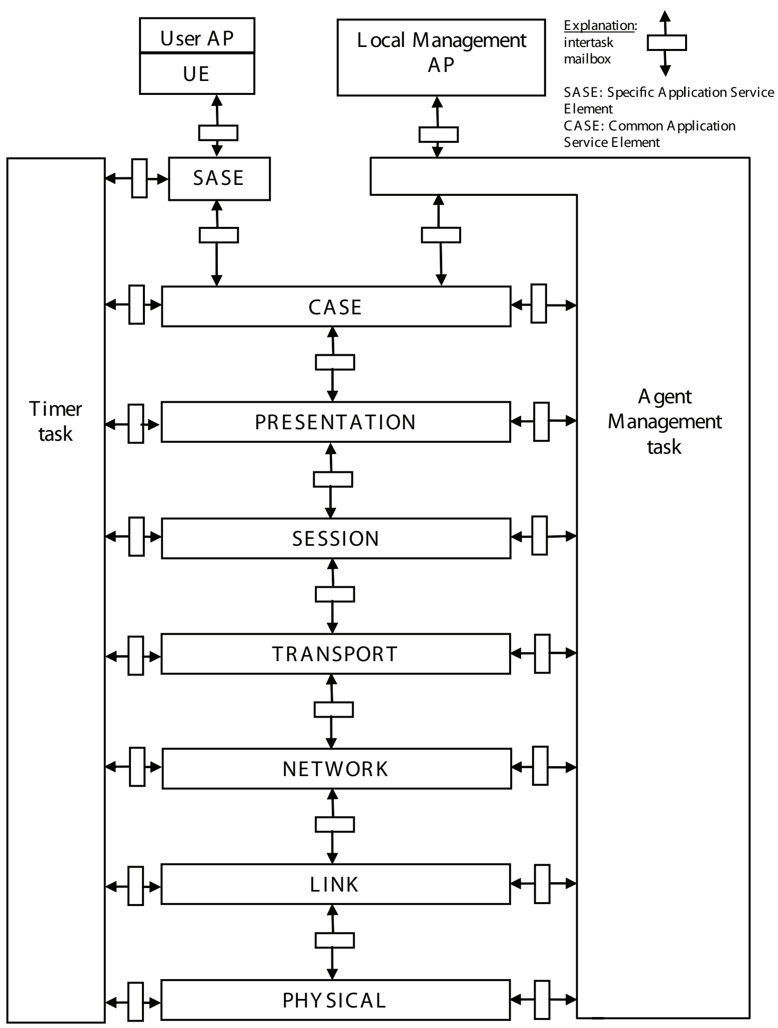
In open systems, three protocol types were created for telecommunications management: system management protocols, layer management protocols and embedded layer management protocols. The most significant system management protocol is the System Management Information Protocol (SMIP). This protocol identifies a system inside a network, also responsible for the activation and deactivation of a system inside of a network, for system configuration and maintenance. The layer management protocols manage a given layer dividing the management into agent tasks and timer tasks. The agent tasks are responsible for parameter setting and error statistics, while the timer tasks monitor the problems related to timeouts, and also provide error statistics reports. The embedded layer management protocols insert management functions into the control field of an existing protocol. A possible control field is a window mechanism field, which is part of the protocol message structure. For instance, a possible window mechanism field contains the sequence numbering, which ensures a network traffic control.

The Common Management Information Protocol (CMIP) is the most significant layer management protocol in the "OSI-World", and is part of the management of network elements (International Telecommunication Union, 1997). The main purpose of this protocol, similar to SNMP, is to support the information exchange among the agents of the network management applications and the managed objects. CMIP was created to eliminate the main problems of the first version of SNMP. Therefore CMIP offers more than SNMPv1 in the following areas: CMIP is more secure: it supports authorization, access control and security logging. In CMIP, the management applications can execute more steps by a single request, and CMIP also offers better notification and report in unusual network operations than SNMPv1. But at the same time, CMIP requires more system resources than SNMPv1 because CMIP exchanges information with agents through managed objects. It sometimes assumes the presence of the Remote Operations Service Element (ROSE) protocol. ROSE is also a layer management protocol, which controls entities to perform remote operations in a distributed environment (see Figure 2).

Management Protocols for Mobile Internet

Nowadays most of the wide area networks are heterogeneous, which means these great networks contain wired and wireless network parts. The management process of the traditional and the mobile Internet networks both require more complex management protocols, because it had become quite difficult to identify users and to approve their rights, so they can access to different services. Authentication, authorization and

Figure 2. Timer task and agent management task.



accounting (commonly used as AAA) represent the 'Big Tree' in terms of IP based network management and policy administration. It collects the resource consumption data for purposes of capacity, trend analysis, cost allocation, auditing and billing.

Authentication provides a vehicle to identify a client that requires access to some system and logically precedes authorization. Authorization follows authentication and entails the process of determining whether the client is allowed to perform and/or request certain tasks or operations. Authorization is therefore at the heart of policy administration. Accounting is the process of measuring resource consumption, allowing monitoring and reporting of events and usage for various purposes including billing, analysis, and ongoing management.

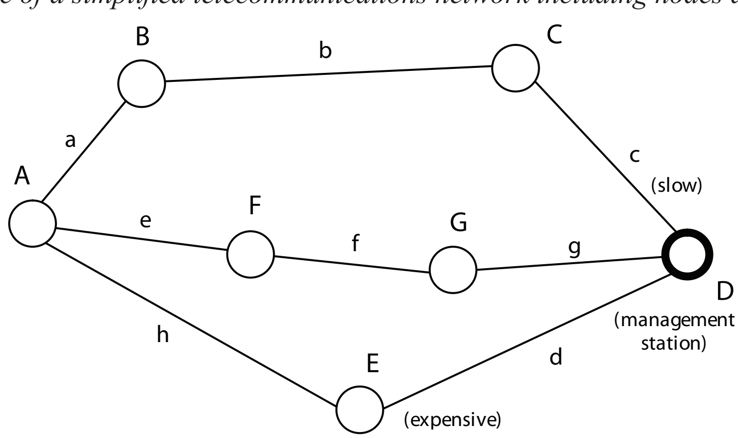
There are four significant AAA protocols: the Simple Network Management Protocol v3 (SNMPv3), the Remote Authentication Dial-in User Service (RADIUS), the Common Open Policy Service (COPS) and the Diameter protocol. SNMPv3 was described earlier at the general presentation of SNMP (Mallenius, 2000).

RADIUS was originally developed for Internet service providers to identify and authorize users. The server running RADIUS not only controls the user authentication, but authorizes and / or rejects the access into the terminal. Additionally RADIUS is widely used by Voice over IP (VoIP) service providers (Hosia, 2003). COPS specifies a simple client/server model for supporting policy control over Quality of Service (QoS) signaling protocols. With COPS, service providers and manufacturers can use a standard data model in setting policy information, dynamically configuring network devices and implementing new services. The Diameter protocol is a successor and upgrade path to RADIUS, being fully compatible with it. Diameter is extensible, has a built-in error notification detection, and is capable to negotiate the capacity of the resources with managed objects.

A SIMPLIFIED EXAMPLE ON TELECOMMUNICATIONS MANAGEMENT FUNCTIONS

Our example presents a telecommunication network which has seven nodes (A, B, C, D, E, F, G). The nodes are connected through channels (a, b, c, d, e, f, g). Both the nodes and the channels have attributes. The attributes of the nodes are the memory capacity and the usage cost, while the attributes of the channels are the signal speed and also the usage cost. In this example, all the nodes have the same memory capacity, but the operation cost of node E is more expensive than

Figure 3. Example of a simplified telecommunications network including nodes and channels



the others. The channels have the same operation cost, but channel c is much slower than the other channels. We take node D as the management station (see Figure 3).

As node D is the management station, the managed objects are nodes A, B, C, E, F, G and channels a, b, c, d, e, f, g, h. The Management Information Base (MIB) is connected to node D. The telecommunication management protocol is composed of the dialogues between the management station and all other nodes. A typical part of a protocol dialogue is the message exchange of request-response messages.

Let us consider two basic scenarios. In the first scenario, the normal operation is described, where we send messages from node A to node D. It is evident that sending the message from node A to node D through node E (via channels h and d) would be the fastest, but because of node E is expensive, we have to consider whether the cost or the speed of the message exchange is more important. In this scenario, we decide to choose a cheaper way for the message exchange. Routing through nodes B and C (via channels a, b and c) or routing through nodes F and G (via channels e, f and g) costs the same, but because of channel c is considered to be slow, therefore we take the best route available, through nodes F and G (see Figure 4).

Figure 4. Example of a simplified telecommunications network including nodes and channels in example scenario number one

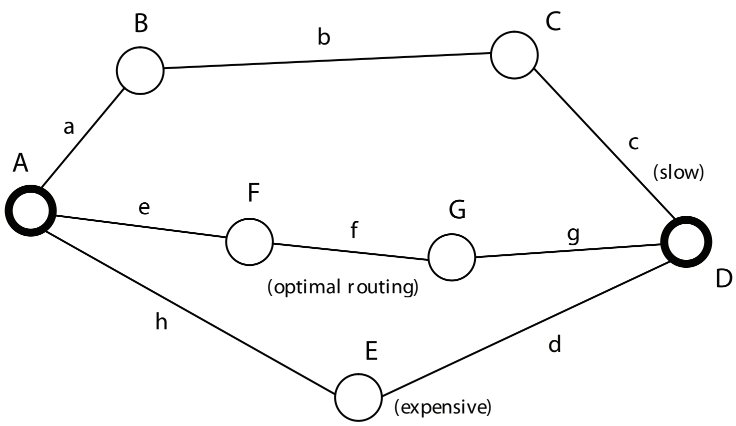
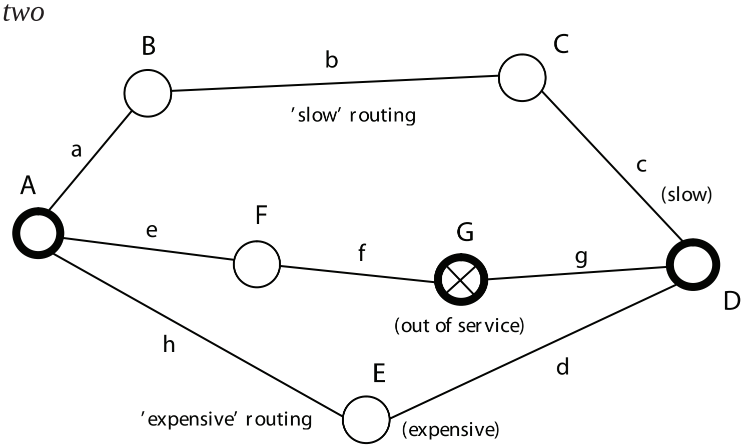


Figure 5. Example of a simplified telecommunications network including nodes and channels in example scenario number two



In the second scenario, we assume that node G is out of service due to a fault. Now we have to find an alternative route considering two available options. If we take the routing through node E, the message exchange will be fast but expensive. If we take the routing through nodes B and C, the message exchange will be slow but still cheaper. After evaluating the requirements of both the user and the service provider, we decide to choose the cheaper routing as the speed of the message exchange is still considered to be adequate (see Figure 5). However, if another channel slows down in this routing (e.g. channel b), we should consider switching to the routing through node E according to the user's requirements.

These two scenarios have presented the fault and the account management. Now we describe how configuration management works in this example. As stated before, there are three parts of the configuration management: name management, data collection and parameter setting. In our example, name management means to list the set of the active nodes and active channels. In the first scenario, all seven nodes and eight channels are operational. In the second scenario, node G is out of service. Because we send messages from node A to node D, generally node F and channels e, f and g are also out of operation.

For instance, data collection deals with timeouts of a request-response message pair, and detects the expected but not arriving response messages. By this operation, this component is also responsible for detecting faults regarding the nodes in the network. In the second scenario, as the message was sent through node G, the undetected response indicated the fault of node G.

In our example, parameter setting is capable for changing the memory capacity of the nodes and the bandwidth of the channels to increase signal speed. By analyzing the signal speed on the routing through nodes B and C, it is optimal to give more bandwidth to channel c to speed up the message exchange if possible.

Now we examine how performance management observes the network operation. Generally, as mentioned before, performance management deals with cost efficiency and general network behavior by creating statistics about the observed events. As an event in the second scenario, node G is out of operation. The performance management now simply collects the response times belong to any request-response pairs, then creates and evaluates the statistical values. From the collected values, a statistical distribution is composed. This is a useful guide for a later parameter setting.

Security management comes last in our example. This function is responsible for authentication, access control and data integrity. In this network, security parameters and security mechanisms should be assigned to network components. We assume that node E has the greatest level of authentication and access control, and channel h has a security mechanism for data integrity. In our example, it means that the messages sent through channel h and node E have the highest level of security.

FUTURE TRENDS

The telecommunications management will continue to evolve in the future. Many new tools and methods will enhance user-friendly services and many technological solutions are expected. The trend is to focus on solutions that satisfy the user's needs and keep the technology hidden. Therefore the applications will become more user-friendly, and at the same time, they have to be more secure. One example of the new ways of security is the identity based cryptography (Housley, 2007).

Furthermore, the future of AAA protocols is pointing towards federations. It means a simpler network access using a new password-based authorization method. For example, users could gain access into a foreign network as guests using their password from their home location (Carlberg & O'Hanlon, 2008).

An important future trend is related to IP-usage and management. The IP Multimedia Subsystem (IMS) is a new framework for delivering Internet protocol multimedia to mobile users. The IMS uses various Internet protocols, such as the Session Initiation Protocol (SIP) developed by the Internet Engineering Task Force (IETF) (Handley, et al., 1999). The trend is to integrate IP-based subsystems into traditional telecommunications systems and networks, because this solution fulfills the user's requirements in using more complex multimedia services. Generally multimedia messaging requires better Quality of Services. To achieve this, the policy-driven management will provide new services to ensure high quality network resources (Haley, et al., 2008; McWalter, et al., 2007).

Most of the management systems have server-client architectures. At this same, there is a trend to establish more peer-to-peer architectures, like the Diameter protocol, which is still under development. Evolution is also expected in mobile location services and new positioning technologies in mobile networking.

CONCLUSION

This chapter gave a brief introduction into the basics of telecommunications management. First the fundamentals and the basic definitions were discussed to give an overview of the managed networks. Next the main management components were presented, starting with the management functions, such as fault management, configuration management, account management, performance management and security management. Then the managed objects were described and represented hierarchically. The Management Information Bases and their data structure were presented, and finally the telecommunications management protocols were classified into three groups: the basic Internet management protocol, SNMP, the basic OSI protocol, CMIP, and the three most significant AAA protocols (RADIUS, COPS and the Diameter protocol). At the end of this chapter, a simplified example was presented and some expected future trends were mentioned.

ACKNOWLEDGMENT

Special thanks to *Dr. Maria Toeroe, PhD* (Senior Researcher, Ericsson Canada, Inc., Montreal) and to *Dr. József Harangozó, PhD* (Associate Professor, Budapest University of Technology and Economics, Budapest) for their useful suggestions. The author is also grateful to *Dr. János Miskolczi, PhD* (Senior Consultant, Ericsson Hungary, Ltd., Budapest), *Tibor Dulai, Dániel Muhi and Szilárd Jaskó* (Assistant Professors, University of Pannonia, Veszprém) for their continuous support and brilliant ideas.

REFERENCES

Buchanan, W. (1999). *Mastering Networks.* London: Macmillan Press Ltd.

Carlberg, K., & O'Hanlon, P. (2008, January). *RFC 5115 Telephony Routing over IP (TRIP).*

Carr, H., & Snyder, C. (2002). *The Management of Telecommunications: Business Solutions to Business Problems.* New York: McGraw-Hill Companies.

Garg, V. K. (2002). *Wireless Network Evolution, 2G to 3G.* New Jersey: Prentice Hall.

Haley, B., Devarappali, V., Deng, H., & Kempf, J. (2008, January). *RFC 5142 Mobility Header Home Agent Switch Message.*

Handley, M., Schulzrinne, H., Schooler, E., & Rosenberg, J. (1999, March). *RFC 2543 SIP: Session Initiation Protocol.*

Hosia, A. (2003). *Comparison between Radius and Diameter (T-110.551 Seminar on Internetwork-*

ing). Helsinki, University of Helsinki, Department of Computer Science.

Housley, R. (2007, November*). RFC 5083 Cryptographic Message Syntax (CMS) Authenticated-Enveloped-Data Content Type*.

International Telecommunication Union (1997). *Information Technology – Open Systems Interconnection – Common Management Information Protocol: Specification.* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.711-199710-I/en

International Telecommunication Union (2000). *Series M: TMN and Network Maintenance: International Transmission Systems, Telephone Circuits, Telegraphy, Facsimile, and Leased Circuits.* Retrieved January 2008 from http://www.itu.int/rec/t-rec-m.3400-200002-I/en

International Telecommunication Union (2002a). *Information Technology – Abstract Syntax Notation One (ASN.1): Specification of Basic Notation.* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.680/en

International Telecommunication Union (2002b). *Information Technology –ASN.1 Encoding Rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER).* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.690/en

Mallenius, S. (2000). *The COPS (Common Open Policy Service) Protocol (Research seminar on IP Quality of Service).* Helsinki, University of Helsinki, Department of Computer Science.

McGinnis, E., & Perkins, D. (1996). *Understanding SNMP MIBs.* New York: Prentice Hall.

McWalter, W., Thaler, D., & Kessler, A. (2007, December). *RFC 5132 IP Multicast MIB.*

Ray, B., & Abbi, R. (2004). RFC 3728 Definition of managed Objects for Very High Speed Digital Subscriber Lines (VDSL).

Stallings, W. (1998). *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2.* Indianapolis, Addison-Wesley Professional.

KEY TERMS

Account Management: Account management enables the measurement of the use of network services and the determination of costs to the service provider and charges to the customer for such use. It also supports the determination of prices for services.

Configuration Management: Configuration management provides functions to exercise control over, identify, collect data from and provide data to network elements.

Fault Management: Fault management is a set of functions which enables the detection, isolation and correction of abnormal operation of the telecommunication network and its environment. The quality assurance measurements for fault management include component measurements for reliability, availability and survivability.

Managed Objects: Managed objects are abstract representations of network resources (a physical entity, a network service, or an abstraction of a resource) that are managed.

Management Information Base: Management Information Base (MIB) is a type of database used to manage the devices in a communications network. It comprises a collection of objects in a (virtual) hierarchical database used to manage entities (such as routers and switches) in a network.

Performance Management: Performance management provides functions to evaluate and report upon the behavior of telecommunication equipment and the effectiveness of the network or network element. Its role is to gather and analyze statistical data for the purpose of monitoring and correcting the behavior and effectiveness of the network, network entities or other equipment and to aid in planning, provisioning, maintenance and the measurement of quality.

Security Management: Security management functionalities include security services for communications (authentication, access control, data confidentiality, data integrity and non-repudiation), security event detection and reporting.

Chapter XXXII
Telecommunications Management Protocols

Katalin Tarnay
University of Pannonia, Hungary

ABSTRACT

In this chapter, the fundamentals of communication protocols are presented, and then a special application area, the telecommunications management is introduced. The main part of this chapter deals with the telecommunications management protocols. First, the model for Telecommunication Management Network is explained, and then the most widely used protocol, the Simple Network Management Protocol (SNMP), is introduced. This is followed by a discussion on the open system management protocols and the mobile Internet management protocols for Authentication, Authorization and Accounting (AAA), their comparison and evaluation. Subsequently, the expected trends are presented. The conclusion part summarizes the content of this chapter emphasizing the main ideas.

INTRODUCTION

The Internet and the Internet services have become an essential part of our everyday life. The importance of IP-telephony, voice over IP and IPTV is growing year by year. Mobility has become a buzzword of Internet services. It is quite difficult to answer whether user mobility, device mobility or vehicle mobility is more important. All of these are supported by different network services which meshes our business and private life. To fulfill all these complex services, the network should be reliable under variable circumstances, and it must be simply available. The users and the service providers equally require high quality services. While the network operations are continuously changing, the Quality of Services (QoS), the time constraints and the simplicity of the network resources should remain unchanged. The telecommunications management itself is responsible for the unchanged network operations under variable circumstances. To summarize, the telecommunications management system monitors, controls and coordinates the network resources, while the telecommunications management protocols are the fundamental tools to fulfill these requirements. The objective of this chapter is to present the main protocols of this field.

This chapter consists of five sections followed by references and key terms. The first section discusses the necessity of telecommunications management protocols and summarizes the structure of them. The next section provides the background information on protocols and telecommunications management, and justifies why this particular method was selected. The next section focuses on the main telecommunications management protocols divided into four parts. The first part explains the Telecommunications Management Network (TMN) model and its architecture. The next part describes the operation of the Simple Network Management Protocol (SNMP), which is the most widely used management protocol. The third part deals with the Open System Interconnection (OSI) protocols, while the fourth part presents the AAA (Authentication, Authorization and Accounting) protocols. The next section forecasts the future trends by elaborating on what kind of new services and protocols are expected. The last section summarizes the main ideas of the chapter. At the end of the chapter, the references and the glossary of key terms are provided.

BACKGROUND

This section provides an overview of the basics of communication protocols and the telecommunications management. The communication protocols can be compared to dialogues. A simple protocol specifies the rules of a message exchange between the network nodes fulfilling some communication task. The rules are applicable to three areas: the message format, the message exchange and the time assumptions. The message is similar to a word in a human dialogue, where the message format is the correctly spelled word. The message exchange means the phrase or sentences told to each other. The correct message exchange has some goal, e.g. inquiring some information. In a network this could be the task of monitoring a network resource. Monitoring needs a request to read a parameter value and a response with the value of that. The response time is limited. It is similar to a human dialogue, i.e. asking a question and waiting for the answer.

The communication protocols are standardized. This worldwide standardization enables the extension of the network. The network nodes have a layered structure, either the five layer of TCP/IP net, or the seven layer architecture according to the Open System Interconnection (OSI) standards. A protocol specifies the message exchange between peer layers of two different nodes. The telecommunications management protocols are generally application layer protocols. A protocol message has typically two parts, a header part and a data part (see Figure 1). Figure 1 shows the general message structure composed of the header and the data. The header contains the message identifier, the identifier of the source and destination nodes. Sometimes an indication of priority is also included. The presence of other fields in the header depends on the function of the message. Some fields contain values that are constant, others values are variable.

The header is followed by the data or information part .The exchange of protocol messages realizes the steps of the protocol operation (see Figures 2 and 3). Figure 2 presents a generic message structure of the message header that includes a message type identifier and a sequence number. The header is followed by information sent to the destination node. Figure 3 illustrates a dialogue

Figure 1. A protocol message.

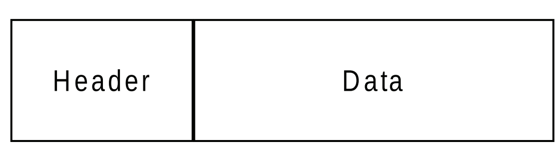


Figure 2. A message example

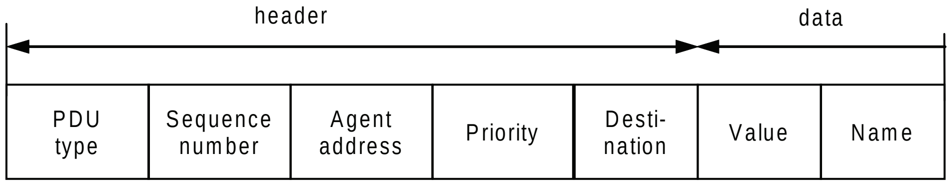
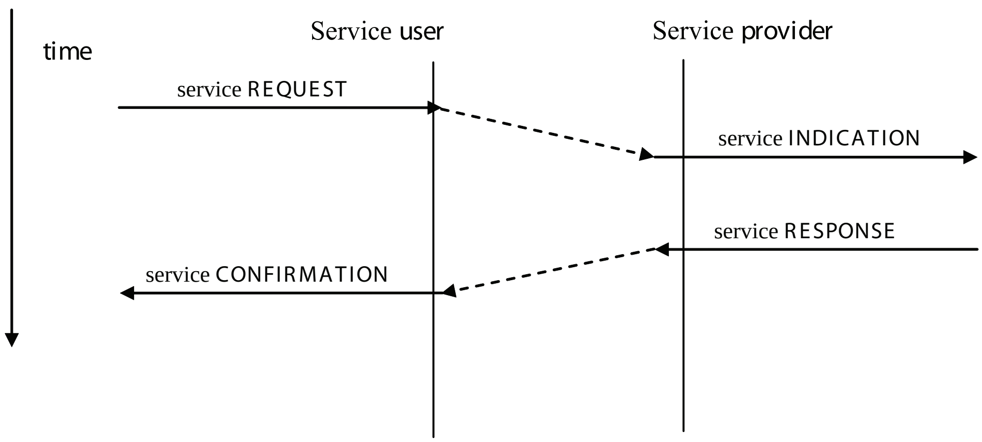


Figure 3. A time sequence diagram: a dialogue between a service user and a service provider



that may take place between a service user and a service provider (Tarnay, 1991; König, 2003).

The main purpose of telecommunications management protocols is the monitoring, controlling and coordinating the network resources to ensure high quality, reliable and error free network operation. The telecommunications management is composed of the telecommunications management station, the Management Information Base, the telecommunications management protocols and the managed objects which are representing network resources. Such a system realizes some management functions. The management functions may be divided into two groups: the basic management functions, which are the fault, account, configuration, performance and security management, and the advanced management functions, which are the authentication, authorization and accounting management. The advanced solutions ensure approximately equal Quality of Services level in the whole network (Garg, 2002; Stallings, 2002). The Quality of Services (QoS) deals with resource reservation rather than the achieved service quality. The packet-switched networks and computer networking in general need protocols supporting QoS. In the field of telephony, an important subset of QoS is the Grade of Services (GoS) which is the ration of lost and sent packets.

While majority of other discussions only deal with one protocol, our presentation familiarizes the reader with all important telecommunications protocols by comparing and evaluating them. The other key feature of our method of discussion is to first of all consider the protocols from the aspects of telecommunications management.

MAIN PROTOCOLS OF TELECOMMUNICATIONS MANAGEMENT

The Telecommunications Management Network (TMN)

The Telecommunications Management Network (TMN) is a protocol model for managing open systems. It is a part of the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) recommendations M-series and is based on OSI management specifications in the ITU-T series X.700. TMN provides a framework for achieving interconnectivity and communication across heterogeneous systems and telecommunication networks. To achieve these goals, TMN defines a set of interface points which perform the communication task (e.g. call processing). The standardized interface element allows the cooperation of elements from different manufacturers using a single management control. TMN can be used in the management of ISDN, B-ISDN, GSM and ATM networks.

TMN has four layers modeling the management of business, service, network, and element. The business management performs functions related to business aspects, analyzes trends and quality. It also provides a basis for billing and other financial reports. The service management performs functions for handling services of the network, like administration and the charging of services. The network management performs functions for distribution of network resources: configuration, control, and supervision of the network. The element management contains functions for handling individual network elements, like alarm management, handling of information, logging and maintenance.

Simple Network Management Protocol (SNMP)

SNMP was the first management protocol used in Internet management. The reason for its rapid spreading was its simplicity. However, this simplicity had also disadvantages, namely it did not support any accounting mechanism. To compensate the lack of these services, the second version of SNMP, SNMPv2 was developed. Currently the most up-to-date version of SNMP is the third version, SNMPv3, which is commonly used in wireless networks.

SNMP is not tied to any particular set of data structures. It refers to a collection of tools used in analyzing network information; the protocol itself is part of this suite. SNMP operates on a collection of related objects identified in a Management Information Base (MIB). Objects in a MIB are identified according to a naming scheme, which is based on Abstract Syntax Notation One (ASN.1) and which uses a hierarchical naming structure. The SNMP model includes a management station, a management agent, a MIB and a network management protocol. In this particular meaning, the management agent is the managed object, however, the terms ‘user agent’ and ‘software agent’ is also often used in telecommunications. Generally, the term ‘agent’ describes a software abstraction, an idea or a concept. The software agent is a piece of a computer program that acts for a user in a relationship of agency, while the user agent is the client application used with a particular network protocol.

In SNMP, the object identifier is used to refer to a single MIB object. The object identifier is a sequence of non-negative integers that represents a traversal of an object tree. This tree starts with the root, while the branches of the object tree are referred to as subordinates. In SNMP, objects are identified by writing the path that leads to get to a specific device.

The protocol operation is divided into two parts: the normal mode and the trap mode. In normal mode the state of managed objects and the parameter values are queried, answered and in some cases, the value is changed. In trap mode the possible errors are detected and collected, and some actions are initialized to inform other management stations to prevent dangerous or even harmful inferences.

During normal operations, the management station sends messages to the management agent, the agent responds to the management station and the management station sends messages to another management station. From a functional perspective, these messages could be reads and writes. Three message types belong to the read function: ‘getRequest’, ‘getNextRequest’ and ‘getBulkRequest’. These message types are sent by the management station to the agent. The agent answers with the ‘getResponse’ message. To write, the ‘setRequest’ message is used by the management station, which is answered in dialogues by ‘getRequest’ and ‘getResponse’ messages.

Read and write messages have well-defined formats composed of fields or group of fields. The first field defined in the message format is the message type. The message types are also referred as Protocol Data Units (PDUs). The next field contains the request identifier, and then the error status and the error index are represented. The variables to be queried, which belong to the ‘getRequest’ and ‘setRequest’ PDUs, are the names and the values. These are enumerated in a predetermined order.

Table 1. Comparison of SNMP versions

	SNMPv1	SNMPv2	SNMPv3
Standards	RFC-1155,1157,1212	RFC-1441,1452 RFC-1909,1910 RFC-1901 to 1908	RFC-1902 to 1908 + 2271 to 2275
Publication year	1988	1993	1998
Structure of Management Information (SMI) - specific data types	simple: integer, octet strings, object IDs application-wide: network addresses (only 32-bit IP addresses), counters (32 bit), gauges, time ticks, opaques, integers, and unsigned integers	New: bit strings Changes: other types of network addresses, counter (64 bit) RFC 1902 (SMIv2)	SNMPv3 uses SMIv2 standard from SNMPv2
Protocol Operations	simple request/response protocol protocol operations: Get, GetNext, Set, and Trap	Similarity: Get, GetNext, Set Changes: Trap message format New protocol operations: GetBulk and Inform	SNMPv3 uses SNMPv2 protocol operations and its PDU message format
Interoperability		SNMPv2 is incompatible with SNMPv1 in two key areas: message formats and protocol operations. Solution: RFC 1908 defines two possible SNMPv1/v2 coexistence strategies: proxy agents and bilingual network-management systems	SNMPv3 is compatible with SNMPv2
Security Features	Community names (Plaintext strings)	SNMPv2: Party-Based security SNMPv2u: User-Based Security (1995) SNMPv2*: User-Based Security (1995) Internet draft only SNMPv2c: Community names (Plaintext strings)	User-Based Security and View-Based Access Control

In trap mode, the protocol operates with trap messages. The first field of the trap message is the PDU type, the second group of fields determines the location the agent (enterprise and agent address), and the third group of fields is the error type, the error code and the timestamp. The last group is the list of names and values of the parameters, called the trap information (Case, et al., 1989; McGinnis & Perkins, 1996; Simoneau, 1999; Zeltserman, 1999).

Table 1 compares the three existing SNMP versions. The first version has a simple, but adequate operation ensuring the 'read' and 'write' functions with 'get', 'getNext' and 'set' messages, and it also supports a simple error detection mechanism with the 'trap' function. Version two is intended to be used in big collections of data in a table format using the 'getBulk' message and for mutual exchange of information between different management stations using the info messages. Version three is similar to version two in its data representation, which is based on structures of management information specific data types (e.g. SNMPv2 has bit strings or counter fields similar to SNMPv3). SNMPv1 only uses a plain text string as a community value. Finally, the higher versions have better security solutions using user-based security.

Open System Interconnection Management Protocols

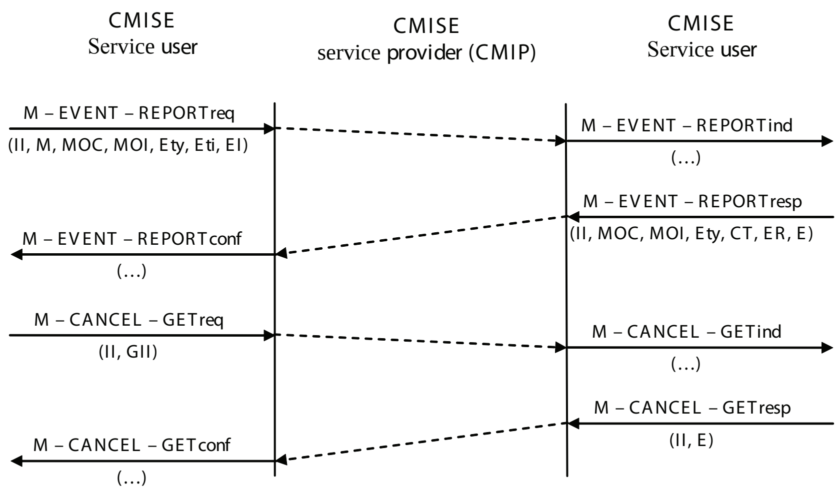
Open System Interconnection (OSI) is a network model where communications are based on seven layers. Each layer has three interfaces: One interface is the protocol which defines the rules of operations between peer network entities, and the other two interfaces are connected to layers above and below through service elements. The protocol is always created in order to provide some communication functions. If the purpose of this communication function is to manage a telecommunications network, then the protocol is called a telecommunications management protocol. The first developed telecommunications management protocol for the OSI network model is the Common Management Information Protocol (CMIP).

CMIP is quite complex, therefore it requires expertise to manage CMIP-based networks. This is the main reason why TCP/IP devices support SNMP instead of CMIP. Although CMIP has been around for more than a decade, it has only a few implementations, mainly in the telecommunications sector. TCP/IP networks and IEEE 802 LANs both have CMIP specifications, described as CMIP over TCP/IP (CMOT), and CMIP over Logical Link Control (CMOL or LAN/MAN Management Protocol, LMMP). In CMIP, similar to SNMP, the management application also exchanges information via managed objects, where these objects are attributes of the managed devices. With the use of the managed objects, these devices can be monitored, controlled and their parameters can be modified.

Implemented by CMIP, defined in the Common Management Information Service Element (CMISE) the following services are available: 'ACTION', to select an action, 'CREATE', to create a copy of a managed object, 'DELETE', to delete a copy of a managed object, 'GET', to query a value of a copy of a managed object, 'CANCEL_GET', to terminate a running query process, and 'SET', to set a value of a copy of a managed object. In the opposite direction, the agent can report an occurrence of an event with the 'EVENT_REPORT' message to the management application (see Figure 4). Figure 4 presents an illustrative dialogue showing how an event report is requested and answered. The time sequence diagram provides a better understanding of the timing of different steps.

To transfer management information between open systems, associations need to be established in the appropriate Open System Interconnection (OSI) layers. Previously the following services were used in the Common Management Information Service to establish associations: 'INI-

Figure 4. A characteristic dialogue as a time sequence diagram.



TIALIZE', to establish association with another CMISE, 'TERMINATE', to terminate an existing association, and 'ABORT', to terminate an association under unusual circumstances. Later, these services were handed over to the Association Control Service Element (ACSE), which is capable of supervising the authentication to establish associations between OSI-method applications. In addition, in CMIP implementations, the Remote Operations Service Element (ROSE) may be used. ROSE is an application layer protocol, which allows entities to perform remote operations in distributed environments.

Table 2 shows the comparison of SNMP and CMIP. CMIP is more secure, but very complex and this has a notable drawback: more complexity means more sophisticated handling. The fundamental difference between SNMP and CMIP is how the devices are defined. CMIP uses an object-oriented method to define the devices individually, while SNMP uses a variable based method, where every variable has its own type identifier. CMIP uses event reports, while SNMP operates with TRAP messages. In conclusion, CMIP offers more opportunities than SNMP, but due to its complex handling methods, today it is rarely used (International Telecommunication Union, 1992; Black, 1995; Raman, 1998).

Table 2. Comparison of SNMP and CMIP.

	Simple Network Management Protocol (SNMP)	Common Management Information Protocol (CMIP)
Standardized organization	IETF	CCITT Recommendation X.711, ISO/IEC 9596
Model	TCP/IP	OSI (CMOT is the variant of CMIP for TCP/IP model)
Security	+	+++
Complexity	+	+++
Protocol level	application	application
Network traffic	-	---

Mobile Internet Management Protocols Remote Authentication Dial-in User Service (RADIUS)

RADIUS (Remote Authentication Dial-In User Service) is an AAA (Authentication, Authorization and Accounting) protocol. Originally it was developed for Internet service providers to identify and authorize their users. The server running RADIUS not only controls the user authentication, but authorizes and / or bans the access to the terminal. RADIUS also logs the events of connection and disconnection of the users, making the protocol suitable for billing purposes. The authentication method depends on the location of the data required. Under simple circumstances the data is stored on the RADIUS server itself. While under ideal circumstances, the information required for authentication is stored on a LDAP (Lightweight Directory Access Protocol) domain controller or on a database server running Simple Query Language (SQL).

Using an access point (AP), the authentication technique is as follows: after the client has reported its intent to connect to the AP, it sends the user name and the password. Until the client has not been identified, it can only use an Extensible Authentication Protocol (EAP). The AP forwards the data (still using EAP) to the RADIUS server. RADIUS sends the login information to the domain controller and also converts the EAP messages to RADIUS format. The domain controller compares the data collected by RADIUS with its own data and responds to RADIUS. Finally, if everything matches RADIUS grants access to the AP.

RADIUS-MIB (RADIUS Management Information Base) is an extension to the MIB used by the Simple Network Management Protocol (SNMP). This fact makes the RADIUS-MIB suitable for the handling of objects to implement the AAA process (as described previously). The RADIUS-MIB has two parts: the client side objects and the server side objects (Rigney, et al., 1997)

Lightweight Directory Access Protocol (LDAP)

LDAP is a directory service mainly used to access a hierarchical database. Unlike the most widely implemented relational databases, it uses a tree graph. LDAP databases are optimized for reading and searching, therefore it is most commonly used when data access and read are more important than data modifications. It can store all kinds of data, as long as they are organized into a tree hierarchy.

Extensible Authentication Protocol (EAP)

EAP is a transfer protocol, a framework for the actual authorization methods. There are five most notable EAP variations.

EAP-MD5 (Message-Digest algorithm 5) supplies the fundamentals of security requirements based on user name and password. It leaves a digital fingerprint on every package it addresses in the data flow. It does not require any kind of client-side Public Key Infrastructure (PKI) certification; therefore the EAP-MD5 is not secure, but it is fast because of its simple structure. This protocol can also be used in wired networks.

EAP-TLS (Transport Layer Security) maintains a high protection level for both the client and the server. The user name and the password are still required, but this protocol also uses a TLS session and requires PKI certifications on both sides. This fact is also the disadvantage of EAP-TLS, because the two-sided certification puts a significant load on the client.

EAP-PEAP (Protected Extensible Authentication Protocol) is also called a protected EAP. The aim of this protocol is to establish a secure transfer agent via an encoded TLS tunnel for the other EAP versions. It does not support authorization by user name and password, but requires confidence between the client and the server in creating and transferring the TLS key. This is a significant difference between EAP-PEAP and the older versions.

EAP-TTLS (Tunneled Transport Layer Security) was developed as an alternative to PEAP. As EAP-PEAP, this protocol also uses an encoded TLS tunnel for the transfer, but clients can also be authenticated by user name and password. EAP-TTLS (unlike EAP-TLS) only requires certification from the server during the authentication process. As this protocol can maintain a high security level without mutual certifications, it is considered to be one of the best solutions for wireless networks.

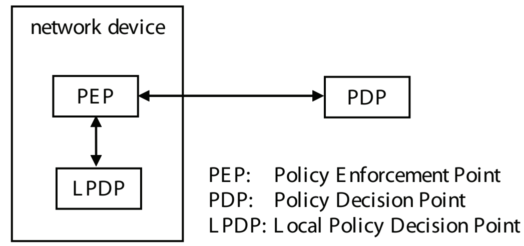
The Lightweight Extensible Authentication Protocol (LEAP) was developed by Cisco and completely proprietary to Cisco, therefore it is only compatible with Cisco products. It requires a two-side authentication: first, the client identifies itself to the authenticator, then the authenticator identifies itself to the client. Only after this procedure can a network access be established. Similar to EAP, the authorization is made by user name and password, and does not require complex PKI certifications. Considering its security level, LEAP is somewhere between EAP-MD5 and EAP-TLS.

Common Open Policy Service (COPS)

COPS is an application layer protocol over TCP (Transmission Control Protocol). In wide area networks, COPS also provides Quality of Services (QoS) at an approximately equal level. From the point of view of COPS, the network has a simple client-server structure, where the client is the Policy Enforcement Point (PEP) and the server is the Policy Decision Point (PDP) (see Figure 5).

The client and the server are connected through COPS, which is used to communicate policy information. The PEP can initiate a connection and sends keep-alive messages in case of no information exchange is presented. If the connection is lost, the PEP tries to establish

Figure 5. The basic model of COPS.



the connection again, if it does not succeed, it tries to reach another PDP. There is a number of COPS messages, for example the ‘In Interface’ message, which defines the source location of the message, the ‘Out Interface’ message, which defines the destination of the message, and the ‘PDP Redirect Address’ message, which defines the next PDP address where the message has to be forwarded in case of former rejection. COPS has two operational models, the outsourcing and the provisioning model. In the outsourcing model, PEP delegates the decision making step to the PDP (Durham, et al., 2000; Mallenius, 2000). In the provisioning model, PEP creates a fixed policy setting before any traffic is initialized. PDP can modify this predefined policy as required.

Diameter Protocol

The Diameter protocol provides the Authentication, Authorization and Accounting (AAA) framework for Mobile IP, delivers attribute value pairs, which carry specific information for AAA, security and configuration. Diameter is extensible through addition of new commands (e.g. can define new accounting or authorizing applications), it has a built-in error notification function and it is capable to negotiate the capacity of the resources with managed objects. Diameter is fully compatible with RADIUS.

A Diameter session consists of commands and exchanges of attribute value pairs between authorized clients and servers. The base protocol provides the minimum requirements for AAA, mobile IPv4 or remote network access applications. In Diameter operation, any node can initiate a request, which makes Diameter a peer-to-peer protocol. The message format includes the message type, the version number and the application type.

Table 3 contains the comparison of the main AAA protocols, SNMPv3, COPS, RADIUS and Diameter. All the four AAA protocols are application layer protocols. SNMPv3, COPS and RADIUS are based on client-server structure, while Diameter has a peer-to-peer architecture. Each of these four protocols has advantages from different points of view. SNMPv3 has the strength of security, COPS can ensure the Quality of Services, while RADIUS and Diameter have good AAA capabilities (Calhoun, et al., 2003; Hosia, 2003; Carnegie Mellon Software Engineering Institute, 2007).

FUTURE TRENDS

In the near future, only small steps are expected in protocol development. To answer questions regarding long term trends is not easy. It seems to be clear that two seemingly different points of view will dominate: the user-centric and the service provider requirements. User-centric solutions ease the user's effort to access and use the network. Service providers and users are both interested in high quality transmission, reliable operation and adequate accounting. The most important requirements are equally important for users and service providers. One of these requirements is security. The telecommunications management protocol messages carry a lot of important information about the message route and its environment; therefore it is a necessity to ensure security. In the near future, the strength of cryptography is expected to be increased, for instance identity based cryptographic methods and new message syntaxes are being developed. As for service providers, routing is an important issue. There are trends dealing with telephony routing over IP, and also the application of IPv6 is under consideration (Boyen, 2007; Devarapalli, 2007; Housley, 2007; Carlberg & O'Hanlon, 2008).

Another interesting new direction is the application of telecommunications management

Table 3. Comparison of AAA protocols

	SNMPv3	COPS	RADIUS	Diameter
Publication year	1998	2000	1997	2003
Main standard	RFC-2271	RFC-2748	RFC-2058	RFC-3588
Standardized organization	IETF	IETF	IETF	IETF
Functions	Collect and set management information on network	exchanging network policy information between a PDP in the PEPs as part of overall Quality of Service	AAA applications	AAA + congestion control
Security	optional IPsec	optional IPsec or TSL	optional IPsec	Always use IPsec or TSL
Transport layer protocol	default UDP, optionally TCP	TCP	UDP	SCTP or TCP
Reliability	+	++	++	+++
Robustness	+	+	++	+++
Architecture	client-server	client-server	client-server	peer-to-peer

methods for other networks. For example, a sensor network is composed of sensors measuring the temperature changes. The management station monitors the temperature in a predefined order and compares the values. This sensor network can be considered as a mobile ad hoc sensor net.

Still, probably the most important future trend is the network convergence, also called the media convergence. It means the coexistence of telephone, video and data communication within a single network. User requirements and consumer demands equally focus on network convergence, especially in the world of the Internet. It is important to note that sending text based information, 'surfing' on the Web, and sending voice messages at the same time demand the significant increase of bandwidth. Also, more sophisticated applications require more effective network resources. Data exchange of increasingly rich contents requires new technologies. Therefore, the network convergence plays an important role in the telecommunications management.

CONCLUSION

This chapter presented the main telecommunications management protocols. First, the general concepts of protocols were briefly discussed, and then the basics of telecommunications management were outlined. The main protocols were classified according to three groups: SNMP, OSI protocols and AAA protocols. The section for future trends was based on ideas for new requirements which should initiate the development of new protocols. The ideas of this chapter can be summarized as: The telecommunications management protocols have a great importance, but despite of them being well developed even today, candidates for new protocols are expected.

The telecommunications management methods will appear in many application areas of e-business. These applications will have an impact on telecommunications. Simplicity, security and openness are the keywords of the future telecommunications management protocols.

ACKNOWLEDGMENT

Special thanks to *Dr. Maria Toeroe, PhD* (Senior Researcher, Ericsson Canada, Inc., Montreal) and to *Dr. József Harangozó, PhD* (Associate Professor, Budapest University of Technology and Economics, Budapest) for their useful suggestions. The author is also grateful to *Dr. János Miskolczi, PhD* (Senior Consultant, Ericsson Hungary, Ltd., Budapest), *Tibor Dulai, Dániel Muhi and Szilárd Jaskó* (Assistant Professors, University of Pannonia, Veszprém) for their continuous support and brilliant ideas.

REFERENCES

Black, U. (1995). *Network Management Standards: SNMP, CMIP, TMN, MIBs, and Object Libraries.* New York: McGraw-Hill Companies.

Boyen, L. M. (2007, December). *RFC 5091 Identity-Based Cryptography Standard (IBCS).*

Calhoun, P., Loughney, J., Guttman, E., Zorn, G., & Arkko J. (2003, September). *RFC 3588 Diameter Base Protocol.*

Carlberg, K., & O'Hanlon, P. (2008, January). *RFC 5115 Telephony Routing over IP (TRIP).*

Carnegie Mellon Software Engineering Institute (2007). *Common Management Information Protocol (CMIP).* Retrieved January 2008, from http://www.sei.cmu.edu/str/descriptions/cmip_body.html

Case, J. D., Fedor, M., Schoffstall, M. L., & Davin, J. (1989, April). RFC 1098 Simple Network Management Protocol (SNMP).

Devarapalli, V. (2007, December). *RFC 5096 Mobile IPv6 Experimental Messages.*

Durham, D., Boyle, E. J., Cohen, R., Herzog, S., Rajan, R., & Sastry, A. (2000, January). *RFC 2748 The COPS (Common Open Policy Service) Protocol*.

Garg, V. K. (2002). *Wireless Network Evolution, 2G to 3G*. New Jersey: Prentice Hall.

Hosia, A. (2003). *Comparison between Radius and Diameter (T-110.551 Seminar on Internetworking).* Helsinki, University of Helsinki, Department of Computer Science.

Housley, R. (2007, November*). RFC 5083 Cryptographic Message Syntax (CMS) Authenticated-Enveloped-Data Content Type*.

International Telecommunication Union (1992). *Management Framework for Open Systems Interconnection (OSI) for CCITT Applications*. Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.700-199209-I/en

König, H. (2003). *Protocol Engineering* (in German). Weisbaden, B. G. Teubner Verlag.

Mallenius, S. (2000). *The COPS (Common Open Policy Service) Protocol (Research seminar on IP Quality of Service).* Helsinki, University of Helsinki, Department of Computer Science.

McGinnis, E., & Perkins, D. (1996). *Understanding SNMP MIBs*. New Jersey, Prentice Hall.

Raman, L. (1998, March*).* OSI Systems and Network Management. *IEEE Personal Communications*, *36*, 46-53.

Rigney, C., Rubens, A., Simpson, W., & Willens, S. (1997, January). *RFC 2058 Remote Authentication Dial In User Service (RADIUS).*

Simoneau, P. (1999). *SNMP Network Management*. New York: McGraw-Hill Companies.

Stallings, W. (2002). *Data and Computer Communications*. New York: Macmillan Publishing Company.

Tarnay, K. (1991). *Protocol Specification and Testing*. New York: Plenum Press.

Zeltserman, D. (1999). *A Practical Guide to SNMPv3 and Network Management*. New Jersey: Prentice Hall.

KEY TERMS

Accounting: Accounting means the tracking of network resources by users. The typical information gathered in the accounting process is the identity of the user, the level of the services delivered, and the time interval in which the services were being used.

Authentication: Authentication means that a user who is requesting services is a valid user of the network services requested.

Authorization: Authorization means that a user can access some services based on its authentication, but the access in only provided for a given amount of time. Therefore authorization is a kind of limitation of the authentication rights.

Outsourced Policy: Outsourced policy is a model where a policy enforcement device issues queries to delegate a decision for a specific policy event to another component external to it.

Policy: Policy is the ability to define conditions for accepting, rejecting and notifying routes according to the actual information.

Provisioned Policy: Provisioned policy is a model where network elements are pre-configured, based on policy, prior to processing event. Provisional policy is contrasted with outsourced policy.

Quality of Services (QoS): Quality of Services refers to an ability to deliver network services according to the parameters specified in a service level. Quality of Services is also an agreement containing service availability, delay, throughput and packet loss ratio, and the ability to provide different priority to different applications, users or data flows.

Chapter XXXIII
Cellular Network Planning:
Evolution from 2G to 4G

Marc St-Hilaire
Carleton University, Canada

Samuel Pierre
École Polytechnique de Montréal, Canada

ABSTRACT

The cellular industry is evolving at a very fast pace. In fact, cellular networks have experienced significant changes over the last few years. In order to keep up with this constant evolution, planning tools must also adapt in order to reflect the particularities and architecture of each generation. In this chapter, we will first present the characteristics and the architecture of the four main generations of cellular networks (1G, 2G, 3G and 4G). We will then expose different planning problems related to each generation followed by a short description of different solutions that have been proposed in the literature.

INTRODUCTION

The cellular industry is a very competitive market in which every service provider wants to maximize their net revenue. A clever way to maximize revenue is to carefully plan the network infrastructure. Due to the extreme complexity of this task, good planning tools are necessary. These tools helped and will continue to guide the network planners in their decision-making process.

The cellular industry is evolving at a very fast pace. In fact, cellular networks have experienced significant changes over the last few years. We saw the first generation (1G) of cellular network in the early 80s followed by a second generation (2G) in the early 90s. A few years later, a third generation (3G) was launched. To date, researchers are already focusing on a fourth generation (4G). In order to keep up with this constant evolution, planning tools must also adapt themselves in order to reflect the particularities and architecture of each generation.

The objective of this chapter is threefold: to present the architecture of each generation of cellular networks; to expose different planning problems related to each generation and finally, to briefly describe different solutions that have been proposed in the literature.

FIRST GENERATION CELLULAR NETWORKS

The first generation of cellular networks appeared in the early 80's. This generation was analog and only used for voice communications. The transmission speed was very limited and the size of the devices was huge compared to what we have today. The following are the main 1G standards.

- Advance Mobile Phone System (AMPS)
- Total Access Communication System (TACS)
- Nordic Mobile Telephone (NMT)

It is important to note that these three standards were not compatible with each other. In fact, mobile users were not able to roam between two different networks thus providing a very limited mobility. A few years later, a second generation was born. That's really with the advent of this generation that the interest started to appear for cellular networks.

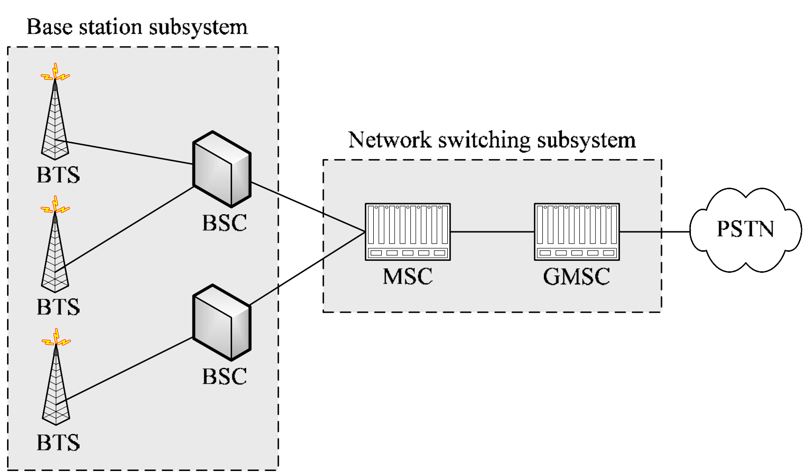
SECOND GENERATION CELLULAR NETWORKS

The advent of 2G networks was a huge milestone since they were digital and capable of carrying voice and data traffic at a maximum speed of 9.6 kbps. Several 2G standards were developed.

- Global System for Mobile Communications (GSM)
- Interim Standard-95 (IS-95)
- Interim Standard-136 (IS-136)

However, the GSM standard was, and still is, the most popular system worldwide. As we can see in Figure 1, GSM networks are composed of two main parts: a Base Station Subsystem (BSS) and a Network Switching Subsystem (NSS). On one side, the BSS is composed of Base Transceiver Stations (BTS) and Base Station Controllers (BSC). The BTS are used to transmit/receive information to/from the mobile users via the air interface. The BSC are mainly used to manage

Figure 1. 2G network architecture



the radio resources and handovers. On the other side, the NSS is composed of Mobile Switching Centers (MSC), often called switches, which are responsible for the switching functions and Gateway MSC (GMSC) to ultimately give access to the Public Switched Telephone Network (PSTN). Finally, different databases such as the Home Location Register (HLR) and the Visitor Location Register (VLR) are also used in order to keep the information related to the subscribers.

Network planners really started to worry about cellular network planning when 2G network were introduced. Different problems such as the coverage, frequency assignment and cells to switches assignment will be exposed in the following sub-sections.

Radio Subsystem Planning Problem

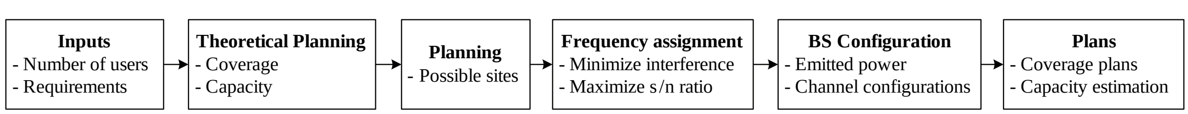
The cost and the complexity of a cellular network mainly depend on the number and the location of the base stations. The radio subsystem planning, also called cell planning, consists to find the optimal location in order to install the base stations. The location selection must comply with the network planner strategy. This strategy can be based on different criteria such as the coverage, the traffic capacity, interference level, handoff, cost, etc. It is important to note that some objectives are contradictory. For example, increasing the traffic capacity might reduce the coverage area or increase the cost of the network if extra base stations are installed.

According to Mishra (2004), the radio network planning process is composed of six steps as shown in Figure 2. The first step consists to collect data. For example, it is really important to know the approximate number of users, the requirements in terms of coverage, capacity and so on. These inputs, in conjunction with empirical propagation models, are then used to make theoretical coverage and capacity plans. Candidate sites are then selected in the next step with the help of engineers. Once the locations have been determined, the following step consists to assign the frequency in order to minimize interferences and/or to maximize the signal to noise ratio (s/n). Then, the base stations must be properly configure (parameter tuning) in terms of power control, handover, etc. Finally, the last step consists of the coverage plans, capacity estimation, budget calculation, etc. It is important to note that radio subsystem planning is a continuous process that continues even after the network is deployed.

2G networks are generally using a Time Division Multiple Access (TDMA) scheme. With this scheme, the radio subsystem planning can be divided in two different sub-problems: coverage and frequency assignment. The coverage planning problem consists to locate the base stations such that the signal will be strong enough to cover the whole area to be planned. To achieve that, link budget calculation and proper propagation models might be required. For the frequency assignment problem, frequencies are assigned to base stations and the goal is to minimize the interference caused by the frequency re-use in neighbors cells. It is good practice to assign different frequencies to neighbors cells in order to have better performance.

Several researchers worked on these two sub-problems. As a result, many mathematical models have been proposed. However, since these problems are NP-hard, approximative algorithms (heuristics) are necessary in order to tackle real size instances of the problem.

Figure 2. Radio planning process



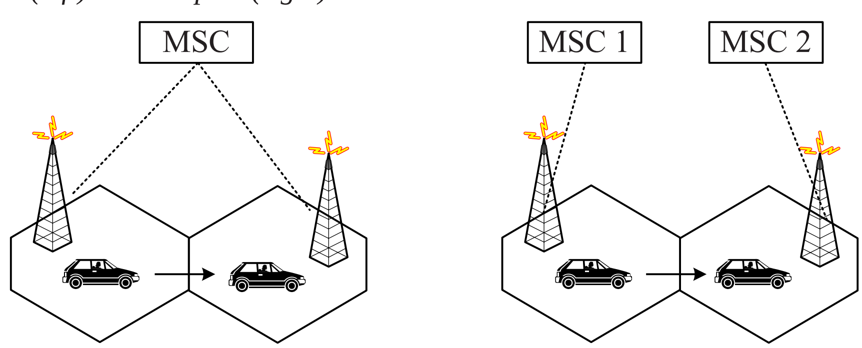
According to Anderson and McGeehan (1994), a good planning tool should consider the propagation environment, the objective function (which is not always uniform) and the possible configuration of the base stations. Keeping these considerations in mind, they proposed an algorithm based on the simulated annealing (see Kirkpatrick *et al.* (1983) for more details about the simulated annealing algorithm) and they used an empirical propagation model. A few years later, Molina *et al.* (2000) introduced a greedy algorithm, a genetic algorithm (see Goldberg (1989) for more details about the genetic algorithm) and a combined algorithm. Their goal was to cover all the clients (also called test points) with the minimum number of base stations. More recently, Hurley (2002) used the simulated annealing algorithm in order to select and configure the base stations. The configuration considers the type of antenna, the power, the tilt and the orientation.

Cells to Switches Assignment Problem

The cells to switches assignment is another interesting problem that can be formulated as follow: Given a set of cells and a set of switches (MSC), how can we assign the cells to the switches in order to minimize a cost function. The cost function is composed of a physical link cost (cost of the links to connect the base stations to the switches) and a virtual handover cost (Pierre and Houeto, 2002). The idea behind this problem is to minimize the number of complex handovers. As shown in Figure 3, a complex handover happens when an active mobile user is crossing the boundary between two cells that are not linked to the same MSC. Complex handovers are much more expensive in terms of resources utilization because several updates are necessary. In order to effectively assign the cells to the switches, we need to know the handoff frequency. If we have a high handoff frequency between two adjacent cells, it would be logical to assign them to the same MSC. This will reduce the resources utilization as well as the operating cost of the network.

The cell to switch problem is well known to be NP-hard (Beaubrun *et al*, 1999). Therefore, approximative methods are also required. Merchant and Sengupta (1995) introduced an integer programming model and a greedy algorithm to solve this problem. A few years later, Saha *et al.* (2000) proposed a few heuristics that are simpler and faster than those previously proposed. In 2002, Pierre and Houeto (2002) proposed a new mathematical equivalence with the p-fixed hub location problem and proposed an approximative algorithm based on the tabu search (see Glover and Laguna (1997) for more details about the tabu search). Several other models have been proposed for this problem. Amoussou *et al.* (2001) used constraint programming to solve the problem while Hedible

Figure 3. Simple (left) and complex (right) handovers



and Pierre (2003) used the genetic algorithm. Finally, multiple homing has been considered by Kubat and MacGregor Smith (2001).

Global Planning Problem

Instead of solving a single subproblem, several researchers tackled more than one subproblems simultaneously. For example, O'Kelly (1987) introduced a quadratic integer program in order to determine the location of the switches and determine the assignment of the cells. Similarly, Klincewick (1991) proposed a heuristic to find the location of the MSC and used different criteria (frequency and distance) to assign the cells. More recently, Cox and Sanchez (2000) introduced a global planning algorithm to minimize the cost of the network. They considered simultaneously the location of the switches, the assignment of the cells and the interconnection (type of links) between the cells and the switches. Finally, Chamberland and Pierre (2002) also considered a global planning approach. They proposed a mathematical model and a tabu search heuristic that consider simultaneously the location and the type of BSC and MSC, the network topology and the interconnection between the network elements.

Reliability Problem

Network reliability can be defined as the capacity of the network to work properly in case of failure. Most of the time, only single failures are considered. A common example can be a base station failure or a link failure between the base station and the switch. Obviously, reliability comes at a price since, most of the time, extra equipments are needed. One common technique to ensure reliable networks is to have nodes and links redundancy. As a result, different algorithms have been developed to plan reliable networks at the lowest possible cost. As we saw previously, Cox and Sanchez (2000) proposed a global planning algorithm. However, they also introduced a new constraint stating that if a node has more than one link going out, it should reach different destination nodes. Therefore, this will provide alternative paths in case of failure. Another interesting paper by Charnsripinyo and Tipper (2003) proposed an overlap by at least two cells at the radio level and a mesh topology at the network level.

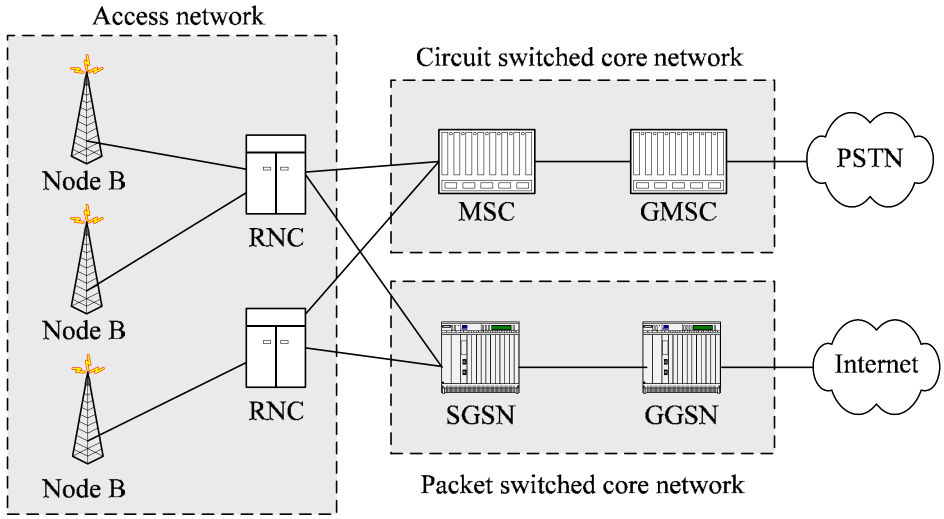
In order to improve the transmission speed of 2G networks, an intermediate generation was developed. This generation is referred to as 2.5G. Two main standards were developed: General Packet Radio System (GPRS) and Enhanced Data rate for Global Evolution (EDGE). GPRS add new equipments in the core network in order to introduce packet switching. Data transfer in GPRS can reach up to 115 kbps. EDGE is an improved version of GPRS and as such, it is sometimes referred to as 2.75G. The main improvements are made at the radio interface and the transmission speed can reach up to 384 kbps. These two standards were only a transition phase before getting to third generation cellular networks.

THIRD GENERATION CELLULAR NETWORKS

Nowadays, third generation (3G) cellular networks are getting out of the lab and start to appear for commercial applications. This latest generation has a greater network capacity (up to 2 Mbps) and has the possibility to offer several value added services (such as multimedia). CDMA-2000 (Code Division Multiple Access 2000) and UMTS (Universal Mobile Telecommunication System) are the two main standards. However, UMTS is actually the 3G standard with the greatest penetration worldwide since it is backward compatible with GSM networks.

As depicted in Figure 4, the UMTS network architecture is composed of two parts: the radio access network and the core network. The radio access network, also called Universal Terrestrial Radio Access Network (UTRAN), is composed

Figure 4. 3G network architecture



of base stations (also called node B) and Radio Network Controllers (RNC). The main functions of the base stations include, but are not limited to, coding, rate adaptation and spreading. The Wide-band Code Division Multiple Access (WCDMA) scheme is used as the air interface between the mobile users and the base stations. The major functions of the RNC (which is an improved version of the BSC used in 2G networks) include load and congestion control of the cell, admission control and routing.

The core network is composed of two domains: the circuit switched core network and the packet switched core network. The circuit switched part inherits from the GSM architecture and is mainly used for real time traffic. It is composed of MSC and GMSC and ultimately provides access to the PSTN. The packet switched part inherits from the GPRS architecture and basically handles all other types of traffic. It is composed of Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) and provides access to the Public Data Network (PDN/Internet).

Radio Subsystem Planning Problem

Since UMTS networks are using WCDMA, the two phases approach mentioned previously (coverage and frequency assignment) is not applicable. In fact, WCDMA shares the bandwidth among all active users such that no frequency assignment is necessary. The capacity of each cell does not depend on the number of connections but on the traffic distribution. In 2G networks, resources were allocated to the users with respect to time and frequency. In 3G networks, users share the power. The power allocated to a given user varies according to its location. That's why a power control mechanism is necessary. For example, a base station can cover a larger/smaller number of mobile users if they are close/far to/from the antenna. As a result, the cell coverage depends on the traffic distribution of each cell. This phenomenon is commonly called cell breathing. When a cell becomes overloaded, the interference will increase and therefore its size will decrease. Excluded users should be covered by less crowded neighborhood cells.

When planning the radio subsystem, it is important to consider the direction of the traffic. In the uplink direction, the traffic is sent from the mobile user to the base station whereas in the downlink direction, the traffic is sent from the base station to the mobile user. It is important to know the direction of the traffic because the constraints are different. For example, when planning the downlink direction, we need to consider the power of the base station and the sensitivity of the mobile user.

Beside the previous mentioned problems, the radio subsystem planning problem also deals with the number, the location and the type of base stations to install. Different aspects such as the cost, the signal quality and the coverage need to be considered. It is common to find papers/models that deal with this subproblem. The majority of these problems require the following as input.

- Set of possible locations to install the base stations
- The location & estimated traffic of each mobile users
- The model and/or the cost of the base stations
- Coverage prediction

Thiel *et al.* (2002) proposed an algorithm in order to find the minimum cost subset of base stations. Starting from an initial solution, they apply a simulated annealing algorithm in order to improve the solution. Finally, local optimizations are made to get a better solution. Amaldi *et al.* (2003a) proposed two greedy algorithms and a tabu search algorithm in order to plan the uplink direction. They consider the installation cost, the signal quality and the traffic coverage. They also considered the downlink direction in a different paper (Amaldi *et al.*, 2003b).

Access Network Planning Problem

The access network is typically made of one or more Radio Network Subsystem (RNS). The latter is composed of one RNC and one or more base stations. The access network basically provides all radio resources and necessary mechanisms to reach the core network. Depending on the assumptions, base stations can be directly connected to the RNC or using a cascaded way. The access network is therefore composed of several trees where the RNC are acting as the root (Harmatos *et al.*, 2000).

In general, the access network planning problem consists to determine the number, the location and the type of RNC as well as the interconnection between the base stations and the RNC. The goal is to minimize the cost.

Several authors focused their research on the access network. Harmatos *et al.* (1999) proposed a heuristic in order to minimize the cost of the access network. Their algorithm optimizes the number and the location of the RNC and the interconnection with the base stations. The cells to switches assignment problem (see section 3.2) was also studied for 3G networks. In fact, Wu and Pierre (2003) proposed a constraint programming algorithm to optimally assign the node B to the RNC. Another approach to assign the node B to the RNC is based on clustering. Lauther *et al.* (2003) proposed two algorithms: a bottom up approach and a top down approach. In the bottom up approach, each node start as its own group and, at each iteration, two groups are joined. You keep merging groups until the cost function can not be reduced. In the top down approach, they first build a tree and then iteratively break the tree into two subtrees.

Core Network Planning Problem

The core network is the center of the network. In fact, equipments installed in the core network must be powerful because almost all the traffic will transit through the core. The core network supports communication with the access network as well as with others core networks.

The core network planning problem consists to install the different equipments (gateways, MSC and SGSN) and to interconnect them to the access network. The goal is to minimize the cost of the core network.

Very few researches have been made on this particular topic. This can be explained by the fact that this problem is similar to fixed (wired) networks. Therefore, algorithms developed for wired networks could also apply to this problem. Harmatos (2002) was the first to present an algorithm especially targeted for the core network of UMTS networks. His algorithm connects the RNC to the core network, finds the optimal locations for the gateways and finally interconnects the transport nodes.

Global Planning Problem

Instead of solving a single subproblem, several researchers focused on more than one simultaneously. Chamberland (2004) tackled the access network and the core network planning problems simultaneously. The objective of his model is to minimize the update cost of the network while considering network performance such as call and handover blocking. A mathematical formulation is presented followed by a tabu search algorithm.

More recently, St-Hilaire *et al.* (2006) proposed an integrated approach in which the three previous subproblems (the radio subsystem planning, the access network planning and the core network planning) are considered simultaneously. This approach has the advantage of providing the optimal cost for the whole network since all the interactions between the subproblems are preserved. The authors introduced a mathematical formulation of the problem as well as a local search and a tabu search algorithms.

In order to improve the speed of 3G networks, an intermediary generation was proposed. This generation is referred to as 3.5G. The two main 3.5G standards are HSDPA (High-Speed Download Packet Access) and HSUPA (High-Speed Uplink Packet Access). The latter is sometimes referred to as 3.75G. Their speed can reach up to 14.4Mbps and 5.7 Mbps respectively. Most of the time, merging from 3G to 3.5G only requires a software upgrade of existing UMTS base stations.

FOURTH GENERATION CELLULAR NETWORKS

Third generation networks are not even fully deployed yet, but we are already researching about next generation networks. Several terms are used in the literature to identify these emerging networks: beyond 3G (B3G), fourth generation (4G) and next generation networks (NGN). Whatever the name used, the key concept of this future generation is integration. In fact, 4G networks will be an integration of different access networks (ad-hoc, sensor, Wi-Fi, Wi-Max, GSM, UMTS, etc.) around a common IP (Internet Protocol) core network. Global roaming across multiple access networks should be possible. Higher bandwidth will also be possible in 4G networks.

One interesting problem in 4G networks is the node-relay placement problem. The latter consists to find the optimal location of relay nodes in order to improve the cellular coverage and provide higher bandwidth. In fact, base stations may not be able to fully cover a given region and provide the requested bandwidth due to the power limitations. Relay nodes are a good alternative since they are smaller and cheaper compared to base stations. Wei *et al.* (2004) and Florea *et al.* (2005) proposed algorithms to solve this problem.

4G networks still bring a lot of challenges since different access technologies will be merged into a single architecture. Here is a list of potential problems for which different solutions will need to be found.

- **Handover between different access networks (vertical handover):** Several mechanisms have been developed for horizontal handovers (within the same access network).

However, how can we execute a handover between two different access networks (from UMTS to Wireless LAN for example)?

- **Accessibility:** How the mobile users can switch among different access networks? Should this be transparent to the users?
- **Quality of Service (QoS):** How to guarantee/maintain an end-to-end QoS across different access networks? How to maintain the QoS requirements during a vertical handover?
- **Security:** How to ensure security if the information is flowing through different networks?
- **Pricing and billing:** Who should be charged for the service received?

Several problems remain unsolved for now. However, 4G networks are still a work in progress. Standards bodies are currently in the process of defining clear recommendations for this future generation.

CONCLUSION

Cellular network planning is a very hot topic. As long as technology will be evolving, new algorithms, methods and tools will need to be developed in order to carefully plan those networks. A proper planning will result in maximizing the net revenue for the service providers and will provide a good quality of service for the mobile users.

Due to the large scale of cellular networks (large coverage area and millions of mobile users), it is very difficult, if not impossible, to develop exact models. For this reason, approximative algorithms (heuristics) are preferred.

REFERENCES

Amaldi, E., Capone, A., & Malucelli, F. (2003a). Planning UMTS Base Station Location: Optimization Models with Power Control and Algorithms. *IEEE Transactions on Wireless Communications, 2*(5), 939-952.

Amaldi, E., Capone, A., & Malucelli, F. (2003b). Optimization models and algorithms for downlink UMTS radio planning. *IEEE Wireless Communications and Networking Conference, 2*, 827-831.

Amoussou, G. E., Pesant, G., & Pierre, S. (2001). Affectation de cellules des commutateurs par programmation par contraintes. *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2*, 1017-1022.

Anderson, H. R., & Mcgeehan (1994). Optimizing microcell base station locations using simulated annealing techniques. *IEEE Vehicular Technology Conference*, (pp. 858-862).

Beaubrun, R., Pierre, S., & Conan, J. (1999). An efficient Method for optimizing the assignment of cells to MSCs in PCS Networks. *Proceedings 11th International Conference On Wireless Communications (Wireless 99), 1*, 259-265.

Chamberland, S. (2004). An efficient heuristic for the expansion problem of cellular wireless networks. *Computers & Operations Research, 31*(11), 1769-1791.

Chamberland, S., & Pierre, S. (2002). An efficient Algorithm for the Design Problem of Cellular Wireless Networks. *10th International Conference on Telecommunication Systems, Modeling and Analysis.*

Charnsripinyo, C., & Tipper, D. (2003). Topological Design of Survivable Wireless Access Networks. *Design of Reliable Communication Networks (DRCN)* (pp. 371-378).

Cox, L. A., & Sanchez, J.R. (2000). Designing least-cost survivable wireless backhaul networks. *Journal of Heuristics, 6*(4), 525-540.

Florea, A., & Yanikomeroglu, H. (2005). On the optimal number of hops in infrastructure-based fixed relay networks. *IEEE Global Telecommunications Conference (GLOBECOM), 6*, 3242-3247.

Glover, F., & Laguna, M. (1997). *Tabu search*. Kluwer Academic Publisher.

Goldberg, D. E. (1989). Genetic algorithm in search, optimization and machine learning. New York: Addison-Wesley.

Harmatos, J. (2002). Planning of UMTS Core Network. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2*, 740-744.

Harmatos, J., Juttner, A., & Szentesi, A. (1999). Cost-based UMTS Transport Network Topology Optimisation. *International Conference on Computer Communication* (I*CCC'99).*

Harmatos, J., Szentesi, A., & Godor, I. (2000). Planning of Tree-Topology UMTS Terrestrial Access Networks. *11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1*, 353–357.

Hedible, C., & Pierre, S. (2003). A Genetic Algorithm for Assigning Cells to Switches in Personal Communication Networks. *IEEE Canadian Review, 44*, 21-24.

Hurley, S. (2002). Planning Effective Cellular Mobile Radio Networks. *IEEE Transactions on Vehicular Technology, 51*(2), 243-253.

Kirkpatrick, S., Gerlatt, C. D. Jr., & Vecchi, M.P. (1983). Optimization by Simulated Annealing. *Science, 220*, 671-680.

Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research, 53,* 25-37.

Kubat, P., & Macgregor Smith, J. (2001). A multi-period network design problem for cellular telecommunications systems. *European Journal of Operational Research*, 439-456.

Lauther, U., Winter, T., & Ziegelmann, M. (2003). Proximity Graph based Clustering Algorithms for Optimized Planning of UMTS Access Network Topologies. *IEEE International Conference on Telecommunications (ICT 2003)*, 1329-1334.

Merchant, A., & Sengupta, B. (1995). Assignment of cells to switches in PCS Networks. *IEEE/ACM Transactions on networking, 3*(5), 521–526.

Mishra, A. R. (2004). Fundamentals of Cellular Network Planning and Optimisation. John Wiley & Sons.

Molina, A., Athanasiadou, G. E., & Nix, A. R. (2000). The automatic location of base-stations for optimized cellular coverage: A new combinatorial approach. *IEEE Vehicular Technology Conference* (pp. 606-610).

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *Journal of Operational Research, 32*, 393-404.

Pierre, S., & Houeto, F. (2002). A tabu search approach for assigning cells to switches in cellular mobile networks. *Computer Communications, 25*, 464-477.

Saha, D., Mukherjee, A., & Bhattacharya, P. S. (2000). A simple heuristic for assignment of cells to switches in a PCS network. *Wireless personal Communications, 12*(3), 209-224.

St-Hilaire, M., Chamberland, S., & Pierre, S. (2006). Uplink UMTS Network Design - an integrated approach. *Computer Networks, 50*, 2747-2761.

Thiel, S. U., Giuliani, P., Ibbetson, L. J., & Lister, D. (2002). An automated UMTS site selection tool. *3G Mobile Communication Technologies* (pp. 69-73).

Wei, H., Ganguly, S., & Izmailov, R. (2004). Ad hoc relay network planning for improving cellular data coverage. *15th IEEE International*

Symposium on Personal, Indoor and Mobile Radio Communications, 2, 769-773.

Wu, Y., & Pierre, S. (2003). Optimization of Access Network Design in 3G Networks. *Canadian conference on electrical and computer engineering (CCECE)* (pp. 781-784).

KEY TERMS

1G: First generation of cellular networks. This analog system appeared in the early 80's and was mainly used for voice communications. AMPS, TACS and NMT are three different 1G standards.

2G: Second generation of cellular networks. They appeared in the early 90's and they were numeric (instead of analog as in 1G networks). Voice and data can now be transmitted at a maximum speed of 9.6 kbps. The main 2G standard is GSM (Global System for Mobile Communications).

3G: Third generation of cellular networks. The goals for this third generation were to reach speed of up to 2 Mbps, provide a worldwide compatibility (global mobility) and be backward compatible. The main 3G standard is UMTS (Universal Mobile Telecommunications System).

4G: Fourth generation of cellular networks. This is the future generation of cellular networks. The requirements for this new generation are not set in stone yet. However, the main goal is to integrate different access networks around a common IP (Internet Protocol) core network.

Handover: When a mobile user with an active call is moving from one cell to another cell (crossing the boundary), the automatic process in which the communication is transferred from the first base station to the second base station is called handover. Simple and complex handovers are possible. In a simple handover, both cells are linked to the same switch (MSC) while in a complex handover, both cells are connected to different switches. Obviously, complex handovers are much more costly in term of resources utilisation since several updates are required.

Horizontal Handover: Handover happening between two cells of the same network. A user moving from a UMTS cell to another UMTS cell is an example of horizontal handover.

Vertical Handover: Handover happening between two different access networks. For example, an active mobile user could be moving from a UMTS cell to a Wi-Fi hot spot.

Network Planning: The goal of network planning is to interconnect different equipments in order to share resources among several users. When planning a network, we want to achieve a given objective. For example, the network planner may want to minimize the cost, minimize the delay, maximize the throughput, etc.

Network Reliability: Capacity of the network to offer the same services even during a failure. Single failures (node or link) are usually considered since they account for the vast majority of failures.

Chapter XXXIV
IP Address Management: Challenges, Solutions and Future Perspectives

Faouzi Kamoun
University of Dubai, UAE

ABSTRACT

IP network connectivity is becoming so important for business that it is being compared today to basic utilities, such as water and electricity. As key enablers of IP network connectivity, IP addresses are positioned today among the most important resources to be managed. Ready access and flawless handling of IP address usage, assignment, tracking and reallocation help in enhancing network reliability and security, while enabling more efficient network expansion and troubleshooting. Today, more than any time before, companies are realizing the need to adopt and implement a comprehensive IP Address Management (IPAM) strategy. Such a strategy, backed by automated IPAM tools, will enable organizations to address the stringent requirements imposed by new converged technologies. This chapter turns the spotlight on the most important challenges in IPAM and attempts to address some of the solutions and best practices to tackle these challenges. The author's views on future IPAM perspectives are discussed and some open research issues are pointed out. The paper underlines the need for organizations to adopt proven IPAM best practices and deploy good automated IPAM tools. This will put them in a better position to expand and leverage their existing networks, while optimizing their IP address space in a secured and controlled fashion.

INTRODUCTION

Most organizations today, whether carriers, service providers or enterprises, rely on IP -based networks to service their information processing and communication needs. Core business applications such as enterprise collaboration, electronic-commerce and internal business operations rely on network integrity, availability, and reliability. IP-centric networks will become even more important with the ongoing integration of converged data, voice and video services over a unified IP-based network infrastructure. The recent proliferation of IP-based devices such as IP phones, wireless PDAs, access devices and RFID readers is further amplifying the role of IP-based networks in enabling next-generation services such as enterprise mobility and mixed-mode unified communications. As a result, IP network connectivity is becoming so vital for business that it is being compared to utilities such as water and electricity (Liu, 2005). As key network connectivity enablers, IP addresses turn out to be one of the most strategic resources to be managed in modern networks.

Broadly speaking, IP Address Management (IPAM) can be defined as an ongoing practice which deals with the proper allocation, assignment, tracking and reallocation of IP addresses. To achieve these tasks, organizations need to formalize and adopt comprehensive IPAM strategies that best suit their needs. As shown in Figure 1, this IPAM strategy spans across three unified management functions, namely IP Address Inventory Management, Domain Name Service (DNS) management and Dynamic Host Configuration Protocol (DHCP) management (Kerravala, 2005). Each of these functions is essential for the proper operation of an IP network, as detailed below.

IP address inventory management provides the ability to remotely handle the allocation of IP addresses within the organization's IP address space. This function is also responsible for archiving IP inventory data and for optimizing usage of the IP space.

DNS management handles the proper configuration of DNS servers, based on best practices. The objective is to keep DNS servers highly secure and available, while keeping DNS configuration data sanitized from syntactical and logical errors ("IP address management demands," 2007).

DHCP management enables the assignment of static and dynamic IP addresses, ensuring that these are available on the fly to authorized hosts that require connection to the network. This requires the proper configuration and provisioning of DHCP servers to maintain their security, integrity, and availability. DHCP management also aims to automate and optimize the usage of the scarce IP address space.

Today, a new breed of IPAM solutions is being proposed, with a promise to automate and optimize the management of the IP address space throughout the IPAM lifecycle, as shown in Figure 2. Some of the strategic tools offered by these IPAM solutions include centralized management and automated provisioning, auditing, diagnostics and reporting. Automation is a key success factor in delivering efficient IPAM solutions that are scalable, flexible and secure and which are capable

Figure 1. IPAM functions and associated four functional areas

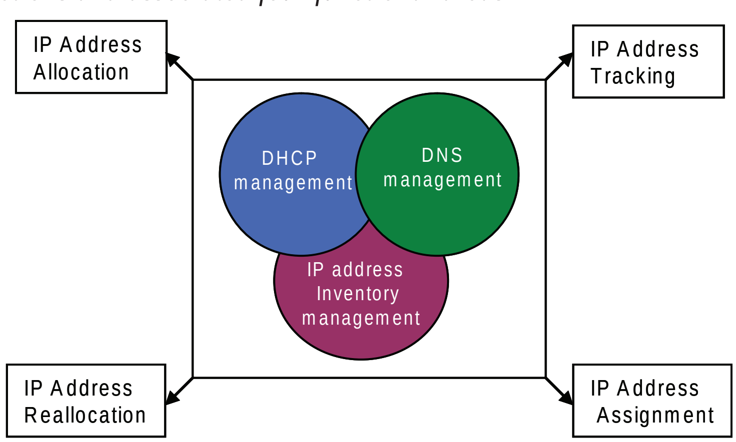
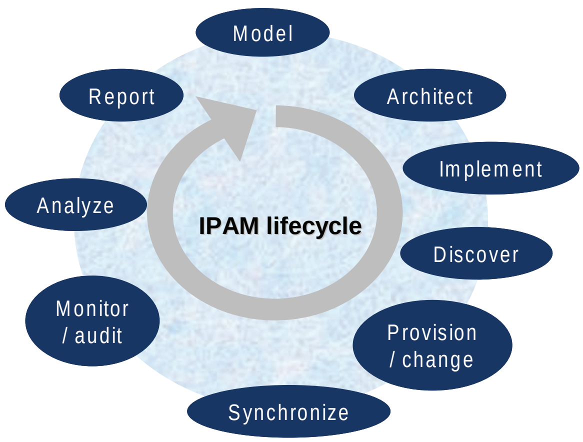


Figure 2. IPAM lifecycle



of addressing the stringent requirements imposed by new converged technologies.

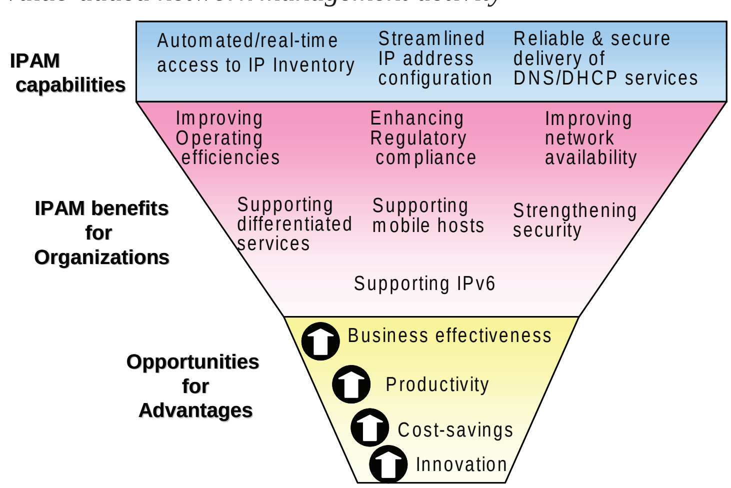
The cost of not adhering to good IPAM strategies and practices can be devastating for organizations. Potential adverse impacts include increased network downtime, revenue losses, decreased IT staff productivity, security breaches, end-user dissatisfaction and slower responsiveness to network changes. In light of the above, organizations need to be aware of the many facets of IPAM, including challenges and best practices.

This chapter starts by highlighting the importance for organizations to adopt comprehensive IPAM strategies, and then summarizes the main approaches to IPAM. This is followed by a discussion of the major IPAM challenges and some corresponding solutions and best practices. Some recent IPAM trends and future perspectives are also explored.

WHY IPAM IS IMPORTANT?

Many organizations have been operating without a comprehensive IPAM strategy, relying mainly on an ad-hoc defensive ("*if it is not broken don't fix it*") approach. Such approaches are typically backed by spreadsheets and very basic homegrown IPAM tools. However the recent advances in network services, protocols, and applications are driving an unprecedented escalation in network complexity. The IP address space is further stretched and exposed, thus requiring more judicious management. As a result, organizations are realizing the need for a unified and holistic approach to IPAM for seamless network management. Some of the benefits of IPAM are highlighted in Figure 3 and are discussed in more detail in the text that follows.

Figure 3. IPAM as a value-added network management activity



Improving Network Availability and Operating Efficiencies

Through its potential to consolidate and automate IP address inventory management, DNS, and DHCP functions, a comprehensive IPAM strategy will not only boost end-users' productivity, but will also improve IT staff productivity. This is reflected in the reduction in IP provisioning errors, problem resolution time, and duplication of efforts. IPAM can also reduce the time spent in auditing and reporting network transactions. It can potentially facilitate the troubleshooting of networks faults and security breaches by keeping up-to-date IP inventory data. These enhanced operating efficiencies lower Total Cost of Ownership (TCO) and bring additional savings in terms of reduced IT staff requirements. This will enable the IT staff to focus on more business-driven tasks (Rainge & Perry, 2004).

Supporting Differentiated Services and Mobile Hosts

Many new services are being rolled out, either as cost-reduction/revenue-generating innovations, or simply to enable organizations to leverage their existing dark fiber infrastructures. This has created an unprecedented growth in the number of network-attached devices and associated services. The integration of a variety of services with different performance requirements necessitates the adoption of sound traffic shaping and policing, priority scheduling, service-level management and pricing policies. These cannot be achieved without a unified and tight management of the IP address space; this is where IPAM comes into play. Another driving force for adopting a comprehensive IPAM approach is the increase in the number of roaming mobile users who need access to the network. By weakening the binding of a given IP address to a specific host, mobility has opened-up new IP address tracking and security management challenges that cannot be circumvented by simple manual IPAM approaches.

Strengthening Security

IP network administrators are also concerned with the erosion of traditional security perimeters due to the proliferation in the number of mobile workers, wireless hotspots and the need to provide suppliers, partners and clients with remote network access (Antonopoulos, 2006). Through better management of the IP address space, IPAM can provide network administrators with prompt access to accurate information related to the usage of IP addresses. A good IPAM solution would also support automated audit trails, event management and network monitoring systems which can help, for instance, to trace suspicious IP traffic and network reconnaissance activities. In short, when combined with firewalls, endpoint access security devices and intrusion detection systems, IPAM can enhance network responsiveness to security threats.

Enhancing Regulatory Compliance

IP address space tracking and control, which were once mandated for large carriers, are now becoming standard practices for enterprises as well ("Proteus," 2007). This recent trend has been partially driven by the increased focus of enterprises on network security and accountability. More importantly, regulatory mandates, such as the Sarbanes-Oxley (SOX) Act and the Health Insurance Portability and Accountability Act (HIPAA), prescribe stricter control over the IP address space utilization. Many regulatory mandates demand that enterprises monitor, log and archive all IP-driven network "transactions", including updates to DNS/DHCP configurations. This might for instance require enterprises to provide audit trails of who had leased a given IP address on the network at a given time. A formal IPAM approach can assist in meeting these regulatory requirements through unified IP address monitoring, auditing and reporting capabilities.

Supporting IPv6

IPv6 expands the 32-bit IPv4 address space to 128 bits, allowing for a 340 x 10^{36} total number of available IP addresses. This exceeds by far the 2^{32} maximum number of available IPv4 addresses. While the notation for an IPv4 address is *x.x.x.x*, where each *x* is represented by a decimal number taking any value between 0 and 255, the IPv6 addressing scheme is based on a more complex notation. Each IPv6 address is written in a colon hexadecimal notation as a string of eight 16-bit hexadecimal numbers (such as D488:3012:0008:0CE1:405B:0000: FFCC: 6B35). Clearly, manual handling of such complex addressing becomes prohibitive, thus making IPv6 strongly dependent on dynamic DNS/DHCP operations and end-station auto-configuration (Antonopoulos, 2006).Through better availability of DNS/DHCP services, a comprehensive IPAM will streamline and automate the management of the more complex IPv6 address space. This will enable better scalability, reduce manual costs and address misconfiguration errors. IPAM would also facilitate the enforcement of IPv6 address allocation and assignment policies, as per RFC 3177, thus facilitating a wider IPv6 adoption. Already this adoption is proliferating across the Asia-Pacific Rim countries due to the rapid growth in the number of Internet connections. The United States Department of Defense (DOD) has mandated network support for IPv6 by 2008. Similarly, the Office of Management and Budget (OMB) requires federal agencies to adopt IPv6 in their networks by 2008. Further, 3G mobile networks, based on release 5 of the WCDMA standards, will be required to adopt IPv6. Rapid adoption of IPv6 is also expected in future due to the widespread use of VoIP, RFID, WiFi, Bluetooth and other ubiquitous devices. This steady increase in the number of network-attached devices is bringing the IP address space to a level that traditional ad-hoc IPAM practices cannot handle.

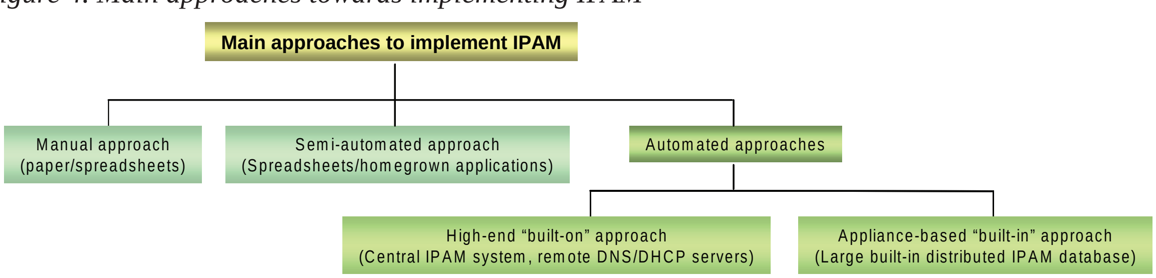
MAIN APPROACHES TOWARDS IMPLEMENTING IPAM

Historically, there have been many approaches for the delivery of IPAM functions, ranging from simple spreadsheets, homegrown tools, to more sophisticated distributed (database-driven) solutions. These approaches can be classified according to the three main categories shown in Figure 4, below.

A manual (paper or spreadsheet-based) approach has been (and is still being) widely used by many small organizations as a cheap, but labor-intensive, alternative to managing static IP addresses. However, to cope with dynamically assigned IP addresses, to accommodate new technologies (such as mobile terminals, VoIP, and RFID) and to quickly adapt to frequent changes in network configurations, automated IPAM approaches become indispensable.

A semi-automated approach relies on the manual configuration of DHCP and DNS servers

Figure 4. Main approaches towards implementing IPAM



to provision the IP configuration data for network devices, while using spreadsheets or homegrown applications to keep track of this information. This approach is criticized for being complicated, prone to human errors and for its weak management and reporting capabilities ("Infoblox," 2007).

An automated IPAM approach has the potential to (1) ease the load on the IT staff, (2) scale-up to support a large IP address space, (3) enable automatic monitoring of dynamically assigned IP addresses, (4) support new IP-based mobile appliances, (5) provide a consolidated repository for real-time access to IP addresses information and (6) integrate full IP address life cycle activities. There are two main approaches to implementing automated IPAM, namely high-end IPAM solutions, which are suitable for very large-scale deployments and appliance-based IPAM solutions, which are suitable for mid-size and large enterprises (Bedard, 2006).

High-end IPAM solutions are typically based on an overlay architecture to manage DNS and DHCP functions. In this model, IPAM functions are delivered by a dedicated IPAM software application, running on a general-purpose computer. This application has its own external database server that is separate from the remote DNS and DHCP servers. The central site generates server configuration and address management data and "pushes" this information to the remote DNS and DHCP servers via agents. This "built-on" approach is often criticized for its high start-up cost as dedicated IPAM software applications and database servers are required (Liu, 2005). Further, additional operational considerations must be taken into account to ensure realtime synchronization of IPAM data across the central IPAM system, the DNS and the DHCP servers. These add-up to the total administration cost of the model. Maintaining a secure and a highly available central IPAM server is another challenge to be carefully considered.

The appliance-based IPAM approach uses multiple distributed and interconnected appliances for the delivery of IPAM, DNS and DHCP services. These dedicated appliances work in concert to create a unified architecture, whereby the database is distributed and embedded in the appliances themselves. Each appliance has full access to the complete database which consolidates network-based DNS, DHCP and IP data (Liu, 2005). Though this integrated "built-in" approach does not generally offer the fully-fledge features offered by high-end IPAM solutions, it does provide a simple, reliable and cost effective mean to implement IPAM. A main challenge to the appliance-based IPAM approach resides in the implementation of a scalable and fault-tolerant distributed database to support the various IPAM functions. Distributed grid management technologies have recently been proposed to address this challenge (see for example Liu, 2005).

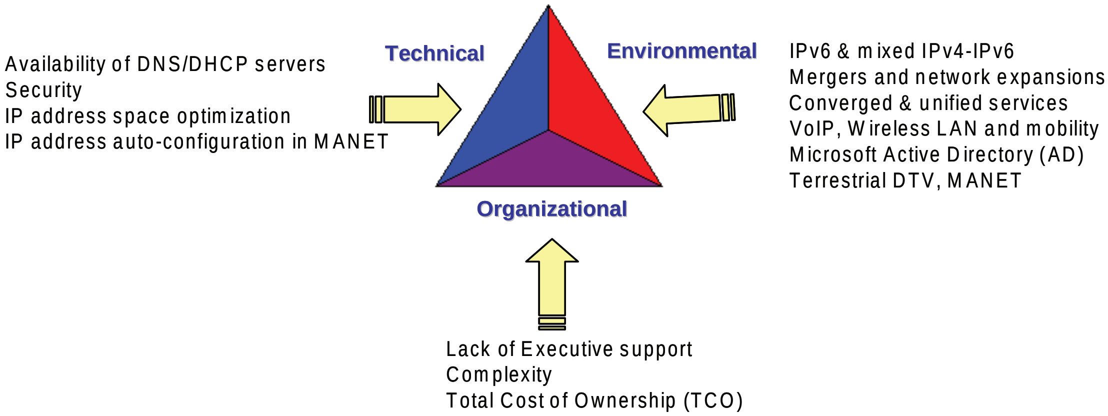
MAIN CHALLENGES AND SOLUTION

While the basic concepts underlying IPAM are relatively simple, the actual implementation can be daunting in the face of the many recent technological challenges. In fact, new requirements have changed the traditional "bookkeeping" role of IPAM, giving rise to a new breed of IPAM strategies and solutions which need to cope with environmental and technical, as well as managerial challenges. Some of these challenges are highlighted in Figure 5 and discussed in more detail below. Solutions and best practices to address some of these challenges are also discussed herein.

Lack of Executive Support

An International Network Services (INS) Industry survey[1] on IP address management revealed that 45% of the surveyed respondents cited the low priority assigned to IPAM projects as the top impediment to improving management of

Figure 5. Main IPAM challenges



IP addresses (Blum, 2007). In fact, compared to other IT projects, IPAM initiatives are still being tagged with the lowest priority in many organizations. This is also reflected in the lack of executive support in endorsing the acquisition of automated IPAM systems and in recruiting qualified staff capable of managing such systems. Addressing this issue can be accomplished by following two main avenues:

First, key providers of IPAM solutions should start serious initiatives to comunicate to CIOs and senior IT decision-makers the value propositions of IPAM. This has more to do with specific business benefits and bottom-line results than tools and technical jargon. For instance, savings from increased IT staff and end-user productivity as well as from increased IT efficiency need to be judiciously quantified to come-up with reliable ROI, NPV and payback figures. Previous successful case studies can also be highlighted to provide further credibility and make stronger impact. For instance Rainge and Perry (2004) interviewed 12 companies in an attempt to quantify the business benefits of automated, end-to-end IPAM solutions. From these interviews, it was found that IPAM enabled companies to achieve an average ROI of 930% with a 106 day payback-period.

Second, as IPAM tools are still being perceived as costly high-end solutions, it is important that IPAM solution vendors offer more economical and scalable IPAM solutions, which are tailored to the various sizes of business organizations. Reducing the investment cost in IPAM solutions will enable higher ROI, while promoting the adoption and proliferation of IPAM solutions among enterprises. To reduce operational costs, IPAM solutions should also be designed to enable ease of use, seamless upgrades, and quick implementation.

Dependency on the Availability of DHCP/DNS Servers

The ability to dynamically manage IP address assignments is strongly coupled with the availability of the DHCP/DNS servers. This challenge can be addressed through proactive monitoring (polling) of DHCP/DNS services, as well as DHCP/DNS failover provisioning. For instance, deploying multiple DNS servers (in master/slave configurations) on different subnets will eliminate single points of failure and enhance DNS service availability. Further, to minimize the chance that the failure of a single DHCP server will bring down all DHCP services, it is important to implement the DHCP configuration according to a distributed and fault-tolerant architecture. Various techniques have been proposed to strengthen the availability and fault tolerance of DHCP services. These include hot standby failover mechanisms, DHCP cluster configurations and split scope methods.

A hot standby mechanism is a traditional failover method whereby two DHCP servers share a common IP address pool. One DHCP server is a primary server in charge of assigning IP addresses, while the other DHCP server is in a hot standby mode, and will only be activated when the primary DHCP server fails.

DHCP clustering implements a local method of DHCP server failover whereby if a failure occurs on a DHCP server (that is a member of the cluster), the server role is transferred to the next server in the assigned nodes list.

In a split-scope configuration, two separate DHCP servers are configured with duplicate scopes, while the scope addresses between the two DHCP servers is divided according to some predefined policies (such as an 80/20 rule). If one server becomes unavailable, the other stand-alone server takes over to lease new IP addresses or renew existing clients.

The ability of DNS/DHCP servers to quickly recover from failures and quickly restart after rebooting is also required to maximize availability. Further, automatic real-time IPAM data backup, restore and disaster recovery can provide seamless protection for the IPAM data. Robustness in the software design of DNS and DHCP servers is also required to ensure that these servers can withstand heavy loads that are of an order of magnitude similar to that generated by Denial of Service (DoS) attacks. Further, non DNS-related traffic which is destined for a DNS server should be blocked to minimize potential threats. It is also good practice to upgrade the DNS and DHCP software applications to the latest releases and to regularly check for the latest security patches and alerts. If dedicated DHCP/DNS appliances are being deployed, then for greater security, it is important to ensure that only a supported, stripped-down version of a hardened OS (such as Linux) is installed. For consistency and uniformity, DHCP/DNS settings, policies and options should be defined, centrally configured and then dispatched to the corresponding servers accordingly.

Coping With Complex and Dynamic Environments

Today's network environment is becoming complex and dynamic due to mobility, network expansions, and the rollout of emerging converged technologies such as wireless LANs, Bluetooth, ENUM, VoIP , RFID, PDAs and ubiquitous appliances. While IPAM strategies and solutions have the potential to address most of the resulting vulnerabilities, there are still many challenges that need to be addressed.

First, mobility entails increased security threats, as DHCP servers are requested to assign dynamic IP addresses to temporary users and mobile resources, while ensuring that these are authorized to access the network. This requires full coordination between the IPAM system and the end-point access control security devices (Antonopoulos, 2006). For instance, an IPAM system can provide security devices with IP address audit trails, which can reveal which IP address was associated with which host at a given time.

Second, as mobility introduces frequent changes to IP address and DNS data, it is important to propagate these changes in real-time across the network. Techniques based on distributed, real-time database and replication server technologies can be used to enable an on-the-fly update of DNS and DHCP data. In addition, periodic and automatic auditing of network IP addresses can pinpoint potential discrepancies between the actual provisioned IP addresses and the corresponding inventory data stored in the DHCP servers (Rooney, 2007). This might occur for instance if a device's IP address has been manually re-provisioned, while evading the regular DHCP-based IP provisioning process. In this case, a network IP address audit would detect and resolve the mismatch.

Third, as wireless technology is opening doors for potential unauthorized access points to the network, it becomes crucial to protect DHCP and DNS servers against malicious attacks (such as DoS and DNS cache poisoning) and unauthorized access. It is also a good practice to automatically monitor requests to access the DNS/DHCP servers, while keeping tight restraint on the Access Control Lists (ACLs) of these servers. Rooney (2007) proposes four different approaches for DHCP servers to identify devices and discriminate IP address assignment. These can be based on MAC addresses, client classes, and user authentication or on an external security scanning system.

Fourth, the recent emergence of new applications such as terrestrial Digital Television (DTV) networks, whereby interactive DTV/IP terminals request IP-based multimedia services from a DTV broadcaster is raising the issue of dynamic IP address set-up for these digital broadcasting terminals. The DHCP/DHCPv6 protocol, which is commonly used for IPAM in Ethernet and WLAN environments, cannot be applied "as-is" because of some unique features of DTV/IP networks. These are described in details by Gardikis, Orfanos, Kormentzas, Pallis and Kourtis (2007). The authors have also initiated the exploration of new IP-based mechanisms for dynamic configuration of terminals in broadcasting networks. They proposed a client-server mechanism, namely, the IP-over-DTV Dynamic Configuration Protocol (IDDCP), which is tailored to the needs of DTV/IP interactive networks.

Management of IPV6 and Mixed IPv4-IPv6 Environments

Though automated IPAM solutions will address many of the issues related to the complexity and length of IPv6 addresses, there are still some challenges related to IPv6 address management. In particular, because of the large installed base of IPv4-based devices, most network implementations in future will support for many years a mix of IPv4 and IPv6 address spaces in a dual stacked environment. Mergers and acquisitions can also lead to mixed IPv4-IPv6 settings. Many transition mechanisms have been proposed to support this co-existence, including configured tunneling, Teredo tunneling, 6to4 automatic tunneling, dual stack, and Intra-Site Automatic Tunnel Addressing Protocol (ISATAP). Handling two parallel IP spaces and the interaction between them requires automated IPAM solutions that can provide centralized management and holistic view to both address spaces simultaneously (Kerravala, 2005). These solutions should also support the various transition mechanisms described earlier and provide unified single-view visualization of the mixed IPv4-IPv6 environment.

Managing Network Mergers and Expansions

Corporate mergers and acquisitions can lead to serious IPAM integration challenges (Bracco, 2004). Besides potential scalability issues, duplicate private IP addresses are likely to surface. The IT staff of the merged organizations has to cooperate and agree on common IPAM conventions, strategy, and tools to manage the IP address space of the merged organization. Many IPAM solutions include tools to facilitate network expansion, without losing the existing DHCP networking configuration. For instance, the "*join/expand networks*" is a distinctive IPAM feature that allows the network administrator to expand a network while preserving the original DHCP configuration, including static addresses and dynamic IP address ranges. In addition, with the anticipated exponential growth of IP-enabled devices, scalability and high performance of DNS, DHCP and IP address management solutions will surface among the top value propositions of IPAM solutions.

Integration With Microsoft Active Directory (AD) Environments

Microsoft's DNS services, which are based on Microsoft Active Directory (AD) technology, are widely deployed for Windows environments today. However in a typical "*heterogeneous*" environment, where different operating systems co-exist, a mix of Microsoft and non-Microsoft DNS servers (such as Berkeley Internet Name Domain (BIND)) is often required. A major requirement for IPAM today is the ability to manage Microsoft DNS servers and interface with Microsoft Active Directory, in a native and non-invasive way, and without the need for external management agents or proxies. This integration of Microsoft with non-Microsoft DNS/DHCP servers poses a challenge (Bracco, 2004). For instance, a need might arise for a non-Microsoft DHCP server to update a Microsoft DNS server. Such operation needs to be executed flawlessly and securely. A single user interface to view multi-vendor DHCP/DNS server configurations will also ease IPAM in large multi-vendor environments.

Optimizing IP Address Space Allocation

To tackle the problem of address space exhaustion, it becomes crucial to optimize the utilization of this space, while satisfying the constraints imposed by the routing topology. This optimization task is best achieved with centralized IPAM tools that have the capability to model a large IP address space hierarchically and then optimally allocate address space to each hierarchy layer (Rooney, 2007). This technique will also facilitate route aggregation for routing protocols and therefore reduces the size of routing tables, while limiting the amount of routing information being advertised.

Automated IPAM tools can also assist in identifying "hotspots" in the network where IP address space is draining. For instance, high watermark thresholds can be provisioned to send alerts to the network administrator when a DHCP IP address range is near exhaustion. This IPAM approach enables the administrator to take proactive measures to expand the range, which will further optimize the usage of the IP space.

Archiving and analyzing historical and current data about IP address utilization statistics across subnets and DHCP pools enable the organization to implement proactive IP address capacity management by predicting trends in address space utilization. Such a forecast can also assist in fine-tuning the usage of the IP space. It can also be used to justify to the applicable Regional Internet Registry (RIR) the need for more space. Today, many IPAM solutions come with unique IPAM modeling features which can assist network administrators to plan for the effective usage of the address space.

IP Address Auto-Configuration in MANETs

A Mobile Ad-hoc Network (MANET) is characterized by a dynamically changing network topology, whereby mobile nodes collaborate to allow communication without the need for a permanent network infrastructure. IP-based routing protocols in MANET rely on each node having a unique IP address, which can frequently change during the node's journey in the network. Some scenarios that can cause a node to change IP address include the merger of two partitions of a network, the merger of two independent MANETs, the merger of a MANET with a LAN, and the usage of a hierarchical MANET addressing scheme (Zhou, Mutka, & Ni, 2004). Furthermore, in a MANET environment, there is no network administrator, nor a centralized control structure. Network partitions and mergers are also frequent, with nodes connecting and disconnecting from the MANET without prior notice. The above unique characteristics make the DNS/DHCP-based IPAM approaches, discussed earlier, impractical in this type of environment. Therefore, new or adapted auto-configuration mechanisms are required.

The challenge of auto-configuring IP addresses in MANETs has been and still is the subject of intensive research studies. Some of the issues of IP address auto-configuration in MANET are discussed in Baccelli, Mase, Ruffino, & Singh, 2007 and include (1) the need for address and prefix generation algorithms capable of supporting operation outside "client-server" frameworks, (2) prefix and address uniqueness requirements and (3) problems related to the management of Internet Configuration Providers (ICPs). IP address auto-configuration schemes can be implemented either in a centralized or in a distributed manner. Centralized schemes (see for example Gunes & Reibel, 2000, Toner & O'Mahoney, 2003 and Sun & Belding-Royer , 2003) are based on a DHCP-like server node which maintains an address pool and which is responsible for address allocation. One disadvantage of this scheme is the heavy processing imposed on the server node, which can also become a single point of failure. Distributed auto-configuration schemes are more

Table 1. Main approaches towards IP address auto-configuration in MANET

	Classification criteria						
Contributors	**Stand-alone MANET**	**Hybrid MANET**	**Protocol independence**	**MANET DAD**	**Partitioning/ merging**	**IPv6 prefix assignment**	**Usage of protocol overhead**
(Perkins, Malinen , Wakikawa, Belding-Royer, & Sun, 2001)	✓	✗	✓	✓	✗	✗	high
(Weniger & Zitterbart, 2002)	✓	✗	✓	✓	✓	✗	high
(Jeong, Park, Kim, Jeong, & Kim, 2006)	✓	✗	✓	✓	✓	✗	high
(Mohsin, & Prakash, 2002)	✓	✗	✓	✗	✓	✗	high
(Tayal, & Patnaik, 2004)	✓	✗	✓	✗	✓	✗	high
(Mase & Adjih, 2006)	✓	✗	✗	✓	✓	✗	none
(Fazio, Villari, & Puliafito, 2005)	✓	✗	✗	✓	✓	✗	medium
(Mase & Weniger, 2006)	✓	✗	✗	✓	✓	✗	none
(Jeong, Kim, Park, & Toh, 2007)	✓	✗	✗	✓	✓	✗	none
(Zhou, Ni, & Mutka 2003)	✓	✗	✓	✗	✓	✓	low
(Ruffino & Stupar, 2006)	✗	✓	✗	✗	✓ (partial)	✗	medium
(Clausen & Baccelli, 2005)	✗	✓	✗	✓	✗	✗	high
(Ros, & Ruiz, 2006)	✓	✓	✗	✓	✗	✓	Medium or high
(Wakikawa, Malinen, Perkins, Nilsson, & Tuominen, 2006)	✗	✓	✓	✓	X	X	Low or high
(Hofmann, 2006)	✗	✓	✓	✓	✗	✓	high
(Adjih, et al., 2005)	✓	✓	✗	✓	✓	✗	low
(Cha, Park, & Kim, 2003)	✗	✓	✓	N/A	✓	✗	medium
(Jelger, Noel, & Fray, 2004)	✗	✓	✓	N/A	✓	✓	Low or medium
(Templin, Russert, & Yi , 2007)	✗	✓	✓	N/A	✓	✓	Low or medium

prevalent and they rely on all nodes cooperating collectively to deliver the functionality of a DHCP server. A good distributed address auto-configuration scheme should be designed to avoid or solve conflicts and duplication among IP addresses, while scaling-up with minimum overhead and low delay/convergence time. The approach should also be designed to support network partitions, mergers and dynamic topologies, while saving energy, memory, and channel bandwidth.

The IETF MANET AUTOCONF working group (Bernardos, Calderon, & Moustafa, 2007) is currently working towards standard specifications and solutions for IP address auto-configuration within different MANET environments. The working group recently conducted a survey on the main proposed IP address auto-configuration approaches. These approaches are classified according to various criteria, such as MANET environment support (stand-alone or hybrid MANETs), Duplicate Address Detection (DAD) support, routing protocol dependency, partitioning/merging support, IPv6 prefix assignment support, and protocol overhead requirement. The main auto-configuration approaches are summarized in Table 1.

FUTURE PERSPECTIVES

As we consider what might revitalize the IPAM landscape in the coming years, we know that quite often the past is a prologue. The growing need to adopt IPv6, rollout next-generation converged and ubiquitous IP services and deliver company-wide end-to-end services will boost the mainstream adoption of comprehensive IPAM strategies and automated tools. The pricing of these tools is expected to decline in the coming years, as IPAM players are under pressure to lower investment costs for potential customers who demand strong IPAM business case with reduced payback and increased ROI. Providers of IPAM solutions should also focus on reducing the prevailing complexity of their solutions. Reducing complexity and lowering prices are key prerequisites to promote the widespread adoption of next-generation IPAM systems. Another existing development is the ongoing efforts of service providers and major IPAM players to educate enterprises on how to move forward with IPv6, VoIP, unified communications and mobility. The rollout of these services will further drive demand for integrated IPAM solutions, making IP inventory management, DNS and DHCP services mission critical to enterprises.

The current trends suggest that the IP/DNS spaces are growing in size, becoming more dynamic, and getting more valuable. These changes are bringing new challenges related to security, scalability, availability, planning, and management. As a result, the role of IPAM will continue to evolve from a simple support and housekeeping function towards a more strategic status, whereby the IP address space is transformed into another strategic asset in the organization.

Next-generation IPAM tools are expected to unleash unique feature sets, based on best design practices. These include:

- Centralized management of IP inventory, DNS and DHCP configurations.
- Real-time IP address repository, with automatic auto-discovery capability and pro-active tracking of IP address space.
- Web-based GUI, which consolidates the view of DHCP, DNS and IPAM information. Hyperlinks and collapsible/expandable tree-based navigation capabilities will provide administrators intuitive information lookup. For large distributed network topologies, new solutions, such as Proteus^{TM} Object tagging, have been proposed to simplify asset tracking with advanced search capabilities, while avoiding navigation through complex IP trees.
- Plug-and-play functionality that reduces installation time, training cost and Total Cost of Ownership (TCO).

- Highly available distributed architecture that can scale up to manage millions of IP addresses, while providing robust failover capabilities.
- Seamless, glitch-free, and remote software upgrades for DNS and DHCP servers.
- Reliance on standard technologies such as XML, SOAP, WSDL, JDBC, ODBC, CORBA, SAX, XAPI and OGSA.
- Support for threshold crossing alerts against IP address space depletion or underutilization.
- Full history tracking and audit trail of address assignments and changes.
- Advanced modeling of the IP address space, based on business policies and automatic mapping of policy-based modeling into DNS and DHCP configurations.

The quest for holistic management of network hardware, software and applications from the Network Operations Center (NOC) will push for the integration of IPAM applications with the Network Management System (NMS). To date, IPAM solutions are delivered as stand-alone applications, and it is expected that these will be integrated in future with the NMS. The NMS handles the main functional areas of network management, including configuration, performance, fault, security, accounting, policy and reporting management. The author believes that IPAM solutions should be packaged within the configuration management framework of the NMS. This framework is already in charge of the auto-discovery and auto-mapping of managed devices, in addition to asset management. The IPAM-NMS integration has the potential to harden network security, while easing fault diagnosis, network monitoring and troubleshooting tasks. However, this integration will not materialize without new initiatives and without the partnership of key providers of IPAM and NMS solutions.

There are several research issues, associated with IPAM in ad-hoc networks. The bounded resources and the dynamicity of the topology call for more efficient IP auto-configuration protocols with low overhead and convergence-time. A related issue arises in the context of network merging where robust duplicate address avoidance and detection algorithms are needed. Further research to explore trust relationship models among nodes participating in the IP auto-configuration process, while protecting against IP spoofing and DoS attacks is also required.

CONCLUSION

The growth in the number of network attached devices, the rollout of new converged technologies, and the proliferation of VoIP and wireless LAN services are creating new and exciting challenges for IPAM. Next generation IPAM solutions need to address the constraints introduced by factors such as low upper management visibility, availability, scalability, security, performance, and TCO. This chapter attempted to provide broader insights to many issues that are hindering the rapid adoption of comprehensive IPAM initiatives. We also discussed some industry best practices and state-of-the-art approaches to adequately address some of the challenges in IPAM. We noted that by adopting proven IPAM best practices and deploying good automated tools, organizations will be ready to expand and leverage their existing networks, while extending and optimizing their IP address space in a secured and controlled fashion.

In future, the anticipated wider adoption of IPv6, along with the expected growth in mobile computing and the rollout of new converged services will make IPAM a mainstream business practice for enterprises and service providers.

REFERENCES

Adjih, C., Boudjit, S., Jacquet, P., Laouiti, A., Muhlethaler, P., & Mase, K. (2005). *Address* *auto-configuration in optimized link state routing protocol*. IETF draft, July 2005, draft-laouiti-manet-olsr-address-autoconf-01.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-laouiti-manet-olsr-address-autoconf-01.

Antonopoulos, M. (2006). *Best practices for securing your IP address management infrastructure*. Special Report, Lucent Technologies. Retrieved November 20, 2006, from http://www.networkworldpartners.com/lucent/pdf/IPAMSRFinal_082906.pdf

Baccelli, E. (2007). *Address auto-configuration for MANET: Terminology and problem statement*. IETF draft, November 2007, draft-ietf-autoconf-statement-02.txt. Retrieved January 4, 2008, from http://tools.ietf.org/id/draft-ietf-autoconf-statement-02

Bedard, D. (2006). *The importance of IP address management*, Vaticor technical article, Retrieved April 6, 2007, from http://searchnetworking.techtarget.com/general/0,295582,sid7_gci1161808,00.html

Bernardos, C., Calderon, M., & Moustafa, H. (2007). *Survey of IP address autoconfiguration mechanisms for MANETs*, IETF draft, July 2007, draft-bernardos-manet-autoconf-survey-01.txt. Retrieved December 27, 2007, from http://ietfreport.isoc.org/idref/draft-bernardos-manet-autoconf-survey/.

Blum, R. (2006). *IP address management*, INS IT Industry Survey, October 2006. Retrieved February 15, 2007, from http://www.ins.com/assets/48615DD3-649A-4ED2-9C62-95196555966B.pdf

Bracco, T. (2004). *Best practices in DNS/DHCP and IP address management*. NetworkWorld special report. Retrieved January 12, 2007 from http://www.genesiscom.ch/de/2/products/downloads/Lucent_IPAM-Bestpractises.pdf.

Cha, H., Park, J., & Kim, H. (2003). *Extended support for global connectivity for IPv6 mobile ad-hoc Networks*. IETF draft, October 2003, draft-cha-manet-extended-support-globalv6-00.txt. Retrieved December 14, 2007, from http://tools.ietf.org/id/draft-cha-manet-extended-support-globalv6-00.

Clausen, T., & Baccelli, E. (2005). *Simple MANET address auto-configuration*. IETF draft, January 2005, draft-clausen-manet-address-autoconf-00.txt. Retrieved December 14, 2007 from http://www.tools.ietf.org/html/draft-clausen-manet-address-autoconf-00.

Fazio, M., Villari, M., & Puliafito, A. (2006). IP address auto-configuration in ad-hoc networks: Design, implementation and measurements, *Computer Networks*, *50*(7), 898-920.

Gardikis, G., Orfanos, S., Kormentzas, G., Pallis, E., & Kourtis, A. (2008). Dynamic IP configuration of terminals in broadcasting networks, *Computer Networks*, *52*(1). 292-302.

Hofmann, P. (2006). *Multi-hop Radio Access Network (MRAN) Protocol Specification*. IETF draft, March 2006, draft-hofmann-autoconf-mran-00.txt. Retrieved December 5, 2007, from http://www.tools.ietf.org/html/draft-hofmann-autoconf-mran-00

Infoblox integrated IP address management solution. (2007). White paper, Infoblox. Retrieved April 12, 2007, from http://forms.infoblox.com/go/infoblox/wpp1001

IP address management demands multi-core architectures. (2006). White paper, BlueCat Networks. Retrieved June 16, 2007, from http://www.bluecatnetworks.com/demo/Proteus_Multi-Core_Architecture.pdf

Jelger, C., Noel, T., & Fray, A. (2004). *Gateway and address auto-configuration for IPv6 ad-hoc networks*, IETF draft, April 2004, draft-jelger-manet-gateway-autoconf-v6-02.txt. Retrieved

December 14, 2007, from http://tools.ietf.org/id/draft-jelger-manet-gateway-autoconf-v6-02.

Jeong, H., Kim, D., Park, J., & Toh, C. (2007). *Passive duplicate address detection for on-demand routing protocols*. IETF draft, April 2007, draft-jeong-autoconf-pdad-on-demand-01.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-jeong-autoconf-pdad-on-demand-01

Jeong, J., Park, J., Kim, H., Jeong, H., & Kim, D. (2006). *Ad-hoc IP address auto-configuration*. IETF draft, January 2006, draft-jeong-adhoc-ip-addr-autoconf-06.txt. Retrieved December 14, 2007, from http://ietfreport.isoc.org/idref/draft-jeong-adhoc-ip-addr-autoconf/.

Kerravala, Z. (2005). *IPv6 success depends on the proper management tools and process*. Enterprise Computing & Networking, Yankee Group Consulting Report. Retrieved September 3, 2007, from http://www.networkworldpartners.com/lucent/pdf/Yankee_IPv6.pdf

Liu, C. (2005). *Build-in vs. built-on IP address management (IPAM)*. White paper, Infoblox. Retrieved June 10, 2006, from http://www.cyberklix.com/PDFs/wp_ipam_05_09_28.pdf

Mase, K., & Adjih, C. (2006). *No Overhead auto-configuration OLSR*. IETF draft, February 2006, draft-mase-manet-autoconf-noaolsr-01.txt. Retrieved December 11, 2007, from http://www.tools.ietf.org/html/draft-mase-manet-autoconf-noaolsr-01.

Mase, K., & Weniger, K. (2006). *PDAD-OLSR: Passive duplicate address detection for OLSR*. IETF draft, June 2006, draft-weniger-autoconf-pdad-olsr-01.txt, Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-weniger-autoconf-pdad-olsr-01.

Mohsin, M., & Prakash, R. (2002, October). IP address assignment in a mobile ad-hoc network, in Proc. *IEEE Military Communications Conference (MILCOM 2002), 2*, 856-861.

Perkins, C., Malinen, J., Wakikawa, R., Belding-Royer, E., & Sun, Y. (2001). *IP address auto-configuration for Ad Hoc Networks*. IETF draft, November 2001, draft-perkins-manet-autoconf-01.txt. Retrieved December 14, 2007, from http://tools.ietf.org/id/draft-perkins-manet-autoconf-01.txt.

Proteus: IP address management built right. (2007). White paper. BlueCat Networks. Retrieved September 14, 2007, from http://download.101com.com/techlibrary/bluecat/bluecat_proteus_ip_address_mgmt.pdf

Rainge, E., & Perry, R. (2004). *Determining the return on investment for automated IP address management*. Whitepaper, IDC. Retrieved May 5, 2007, from http://www.networkworldpartners.com/lucent/pdf/LucentVitalQIPROIWP_Final.pdf

Rooney, T. (2007). *Best practices for next-generation IP address management*. White paper. BT Diamond IP. Retrieved October 24, 2007, from, http://www.pmi.it/file/whitepaper/000133.pdf

Ros, F., & Ruiz, P. (2006). *Extensible MANET auto-configuration protocol (EMAP)*. IETF draft, March 2006, draft-ros-autoconf-emap-02.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-ros-autoconf-emap-02.

Ruffino, S., & Stupar, P. (2006). *Automatic configuration of IPv6 addresses for MANET with multiple gateways (AMG)*. IETF draft, June 2006, draft-ruffino-manet-autoconf-multigw-03.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-ruffino-manet-autoconf-multigw-03

Tayal, A., & Patnaik, M. (2004). An address assignment for the automatic configuration of mobile ad hoc networks, *Personal and Ubiquitous Computing, 8*(1), 47-54.

Templin, F., Russert, S., & Yi, S. (2007). *MANET auto-configuration*. IETF draft, March 2007, draft-templin-autoconf-dhcp-07.txt. Retrieved December 20, 2007, from, http://tools.ietf.org/id/draft-templin-autoconf-dhcp-07.

Wakikawa, R., Malinen, J., Perkins, C., Nilsson, A., & Tuominen, A. (2006). *Global connectivity for IPv6 mobile Ad-hoc networks.* IETF draft, March 2006, draft-wakikawa-manet-globalv6-05.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-wakikawa-manet-globalv6-05

Weniger, K., & Zitterbart, M. (2002). IPv6 autoconfiguration in large scale mobile Ad-hoc networks, in Proc. of *European Wireless 2002* (pp. 142-148). Florence, Italy, February 2002.

Zhou, H., Mutka, M., & Ni, L.M. (2004). IP address handoff in MANET. *Computer Communications, 23*(1), 2425-2433.

Zhou, H., Ni, L. M., & Mutka, M. W. (2003). Prophet address allocation for large scale MANETs. *Ad Hoc Networks Journal, 1*(4), 423-434.

KEY TERMS

Audit Trail: A chronological sequence of logs and audit records, each containing evidence of completed or attempted transactions (such as system activities, operations, events, user access requests, and services). Audit trails are used to reconstruct or investigate user actions or sequence of events that are behind an operation or a system response.

DAD (Duplicate Address Detection): A feature of the IPv6 networking stack that can detect the presence of a duplicate address that is already being used on the network.

DHCP (Dynamic Host Configuration Protocol): An application layer protocol in the TCP/IP suite that automates the assignment of IP addresses, subnet masks, default gateway, and other IP parameters on a network.

DNS (Domain Name System or Service): An Internet service that translates human-readable computer hostnames into IP addresses. DNS servers maintain a distributed database of numeric IP addresses and their corresponding domain names and resolve them to locate remote machines.

IPAM (IP Address Management): A framework designed to simplify the planning, tracking, and management of the IP address space, while managing dynamic IP address services (DHCP) and IP name services (DNS).

NMS (Network Management System): A platform that contains a set of applications which reside on a management station. The NMS is used to remotely monitor, control and administer a network.

RIR (Regional Internet Registry): A non-profit agency that manages the distribution and registration of Internet number resources (IP addresses and BGP autonomous system numbers) within a particular region of the world. There are currently 5 RIRs in operation: ARIN for North America, RIPE NCC for Europe, Middle East and Central Asia, APNIC for Asia and the Pacific region, LACNIC for Latin America and the Caribbean region and AfriNIC for Africa.

TCO (Total Cost of Ownership): A calculation that assesses the sum of all costs of an asset. This includes the cost to acquire, deploy, operate, upgrade and maintain an asset. Technical support and training costs are also included in the TCO.

ENDNOTES

1 This web-based survey, conducted by INS in 2006, was completed by 125 IT professionals around the globe.

Chapter XXXV
Digital Cable TV Networks:
Converging Technologies, Value-Added Services and Business Strategies

Ran Wei
University of South Carolina, USA

Zizhong Zhao
Communication University of China, China

ABSTRACT

This chapter focuses on digital cable TV networks as a convergent network with telecommunications networks and the Internet that provides broadcasting TV and radio, telecommunications services, and IP-based publishing and e-commerce. The chapter first traces the technological evolution of cable TV, highlighting recent developments in digitalization and convergence. The transformation of cable TV networks from channel operators to unified platforms is discussed. In doing so, the key terms and concepts in cable TV technology are introduced. The technological, political, regulatory, and economic forces behind the convergence are also identified. Furthermore, this chapter examines the value chain and collaborative opportunities among the participants in the digital cable TV revolution. User-centered business models of managing digital cable TV networks are proposed.

INTRODUCTION

This chapter provides an overview of the on-going transformation of cable TV networks. As a result of technological advances, cable TV networks are transforming themselves from analog broadcast networks into digital cable networks of broadcast, telecommunications, and broadband Internet, so-called "triple-play services" networks. The next-generation of digital cable TV networks will be a unified platform for delivering hundreds of TV channels, numerous add-on services, telephony services, and the Internet over fiber optic cables. The chapter focuses on tracing the historical evolution of cable TV networks, analyzing the role of new technologies in their transformation, and discussing opportunities and challenges in overhauling the existing cable TV infrastructure.

The key concepts are cable TV networks, digital conversion, video compression technologies, digital cable TV broadcast, media convergence, and two-way cable networks as a multiplatform. They provide the necessary theoretical ground for a thorough understanding of the digitization process and the convergence of broadcast, cable, telephone, and the World Wide Web. Moreover, the chapter will offer new models in managing digital cable TV, which are in a position to offer a wide range of new services, content, and applications. The chapter ends with a brief presentation of challenging issues and management tools.

THE EVOLUTION OF CABLE TV NETWORKS

This section traces the origin of cable TV networks, including the concept of CATV, cable network infrastructure, basic cable technologies, and the major transformations of cable TV networks. Recent developments in upgrading cable TV networks are highlighted; they include the deployment of fiber optics and digitization. The on-going technological upgrades empowered cable TV networks to enhance interactivity and paved the way for them to provide a huge number of channels and add-on services.

The Infrastructure of Cable TV Networks

Cable TV networks refer to a system of transmitting and distributing broadcast signals to TV households using a network of fixed cables (Harte, 2007). Cable TV networks have a humble origin. They began as nothing more than a simple community-based distribution system for receiving clearer broadcast TV signals in remote mountainous areas in 1949. The system included large community antennas and coaxial cables, which distributed the signal received from the antennas to individual households. It was thus known as "Community Antenna Television" or "CATV" (Guillory, 2006, p. 71).

Since the 1950s, cable technology was used in other areas in the U.S. It was during this period, CATV expanded into cable TV networks, which offered more over-the-air TV channels by importing distant signals. This expansion was characterized as the first incarnation of cable TV (Bates & Chambers, 2004). The year of 1964 was a milestone in the development of cable TV networks because American TV households subscribing to cable TV reached one million. After reaching this milestone, the growth in subscribers was gradual but substantial, totaling 41.2 million in 1987 (NCTA, 2007). In 1990, cable TV had reached more than half of the American households with a penetration rate of 50.5%. The rapid growth of cable TV networks during the 1970s and 1980s was largely due to the fact that subscribers were able to view more channels and had an enriched viewing experience. The launching of cable-only channels such as Home Box Office (HBO), EPSN, and Cable News Network (CNN) as well as the availability of super-stations such as WTBS and WGN on cable systems, contributed to the growth of cable TV networks.

As more and more TV channels were launched on cable TV networks, the channel-carrying capacity of cable systems in the United States increased significantly. Early systems carried a total of 33 channels, which tripled to 90-plus channels. They became a viable competitor with over-the-air TV networks. Multichannel cable networks were described as "television of abundance" (Bates & Chambers, 2004, p. 178). This transition was described as the second incarnation of cable TV (Bates & Chambers, 2004, p. 173). Cable TV networks completed the evolution from CATV to broadcast media.

Technically, traditional cable TV networks are analog systems. They include five components: (1) a headend (the facility used by cable service providers to receive broadcast signals from satellites or other means such as microwave links), (2) trunk cable (it distributes the signal from the headend to communities), (3) distributor cables, also called feeder cables (they feed the signal into communities passed by trunk cable), (4) subscriber cable (it delivers the signal to individual homes), and (5) terminal equipment (TV sets that are cable ready; set-top boxes would be provided by cable service operators for sets that are not cable ready)(Cicora, Framer, & Large, 1999).

The growth of cable TV networks in the 1990s was slow as compared to the peak years in growth between 1975 and 1985. The growth in the 21st century was the slowest. According to Nielsen Media Research (2006), the total number of cable households in the United States as a percentage of all TV household fluctuated between 68% and 70% between 2001 and 2005. The annual increase rate was 2%. The small gain in subscribers raised the issue of cable saturation in the country. The growth of cable households in the rest of the world, however, has been robust since the 1990s.

Coaxial cables that are used to carry broadcast TV signals typically consist of two separate wires: (1) a copper wire protected by an insulator, and (2) a wire surrounded by a metallic cylinder. The two-wire cable was designed to reduce electrical interference. However, analog cable TV networks built on metal wires have limitations. Their channel-carrying capability is limited due to small bandwidth (the maximum reached 90 channels in 1998). In addition, broadcast signals need to be strengthened by a set of amplifiers. Otherwise they would become weak and delayed. When this happens, subscribers see double image on TV screen (commonly known as "ghosting").

The Overhaul of Analog Cable TV Networks

During the 1990s, there were two significant technological updates of cable TV networks (Collette, 2004) with a price tag of $50 billion (Bates & Chambers, 2004): (1) replacing coaxial cables with fiber optic cables, and (2) the conversion from analog to digital format in distributing broadcast signals.

First, cable TV networks started to replace metal wires with fiber optic cables for the trunk cables carrying broadcast signals from the headend to communities. Though fiber optic cable system debuted in 1976, it was in the past ten years that cable systems upgrade took place on a massive scale. The network overhaul is on-going. Given the large capital required to overhaul the analog networks, cables TV system operations have to do it gradually. Some of them now operate on a hybrid network built on both coaxial and fiber optic cables (Harte, 2007).

Fiber optics is thin strands of glass; their diameter is thinner than human hair. Normally, hundreds or even thousands of fiber optics are bundled into cables to transmit broadcast signals. Compared to coaxial cable TV networks, fiber optic cable TV networks have several advantages. For example:

1. They are cost-effective in carrying broadcast signals because it costs much less to manufacture them. Also, copper wires distort signals. Amplifiers are used. There is less signal attenuation over fiber optic networks. Hence, the need for building a set of amplifiers on cable TV networks is eliminated, cutting down cost in building and maintaining the cable infrastructure.
2. They improve reception quality thanks to less signal degradation along the fiber optic cable. This is possible because signals transmitted on fiber optics do not interfere with each other, even though they are bundled in the same cable;
3. Because fiber optics is thinner than copper wires, more of them can be bundled into a cable, allowing more channels to be delivered to subscribers. The increase of available transmission bandwidth per fiber

is even faster than copper wires; fiber optic cables can carry multiple channels up to thousands. The age of 500 TV channels, a future envisioned by TCI's John Malone in the early 1990's, has arrived.

4. Although copper cables can also carry digital signals, fiber optics cables are most suited for carrying digital signals, regardless data or telephone calls. On a typical fiber optic cable TV network, digitized information ranging from entertainment, distance education to highly detailed images for remote medical diagnosis can be transmitted. As a computer network carries only digital signals, this technological attribute paves the way for the convergence of cable TV networks with the Internet.

Second, cable TV networks started the conversion from analog video to digital broadcasting. The conversion has reinvented cable TV networks from a network distributing a relatively large number of broadcast TV channels to "a broadband interactive telecommunications network" (Bates & Chambers, 2004, p. 173).

Simply put, digital is a binary system that presents information as either 1 or 0. Thus, digitization refers to the process of converting information (no matter it is in the form of text, graphics, image, music, voice, video or data) into a series of zeroes and ones. When this happens, the information is in the digital format. The result of digitalization is high-quality reproduction of the information, which can be electronically transmitted, manipulated, and stored.

In addition, digital cable TV networks rely on video signal compression technologies to transmit digital information and data. Compression refers to reduction of the quantity of digital data used to represent video images (Symes, 1998). The most widely used compression standards were developed by MPEG (Moving Picture Experts Group). The most widely used standard for digital TV is MPEG-2. This is the reason why the digital cable TV network is also defined as a cable TV distribution network using digitally compressed video.

In sum, digitization and video compression technologies significantly increase the variety of programming available and distributed on cable TV networks. Digital cable services worldwide are rapidly expanding to include add-on interactive services such as on-demand programming. Moreover, digital cable TV networks are inherently two-way modes (download and upload) as compared to analog cable TV networks which support one-way mode (download only). Thus, digital cable TV networks have become a new interactive platform for delivering hundreds of TV channels and customized add-on services.

Expanding Services and New Applications of Digital Cable TV Networks

Thus, digital cable TV networks have the capacity to run an enormous number of channels. Subscribers can view channels devoted to almost everything from 24-hour news, movies, cooking shows, sports, and music to home decorating shows. More important, cable TV networks are transforming themselves from channel operators to an information gateway with a unified platform. The platform is capable of delivering high-definition TV and advanced digital video, audio, and data services on multiple channels (Harte, 2007). With a digital set-top box (STB) and on-screen TV guides, cable subscribers can get such interactive services as video on demand (VOD), pay per view (PPV), near video on demand (NVOD), and personal video recorders (PVRs). These new services are delivered at high speed and allow personalization of content and services (Damascio & Ferreira, 2004; Harte, 2007).

Moreover, digital cable TV networks are capable of providing telecommunications services, including the standard voice over Internet protocol (VOIP) systems. The digitally upgraded network

Table 1. Comparing analog cable TV networks with digital cable TV networks

	Analog Cable TV Networks	Digital Cable TV Networks
Communication Model	One-to-many, mass communication	Point-to-point and one-to-may interactive communication
Network Type	Coaxial cables, wired only	Fiber optic cables, wired and wireless
Headend connectivity	Large antennas	Satellites and microwave links
Channel capacity	Limited	Almost unlimited
Subscriber access	Converter box	Digital STB
Terminal Device	TV	TV with PC-like interface
User Interactive Device	Remote control	EPG
Interactivity	Mainly one-way mode	Two-way mode
Metaphor	Community Antenna TV	All-in-one Network; Web Portal on TV screen
Services	TV	TV, telecommunications, Internet broadband, and the Worldwide Web

also enabled cable companies to offer high-speed Internet access. In fact, the growth of Internet connection services via cable modems has been faster than digital subscriber line (DSL) services offered by telephone companies. Homes passed by cable high-speed Internet services reached 117.7 million in 2007 (NCTA, 2007). Running at a speed up to five million bits per second (mbps), cable modems are much faster than dial-up connections. A range of e-commerce applications, such as online banking and shopping, is now available on digital cable TV networks. Table 1 compares the major differences between traditional cable TV networks and digital cable TV networks.

THE IMPACT OF CONVERGENCE ON DIGITAL CABLE TV NETWORKS

As the growth rate of analog cable TV households has been flat, the future of cable TV networks will depend on digital technologies that enable service providers to launch interactive TV programming and added-on services. This section analyses the impact of convergence on digital cable TV networks, which have expanded to include voice, data, and other multimedia content. It also identifies the enabling factors that facilitate the growth of digital cable TV networks with a unified platform.

Convergence and Factors Facilitating Media Convergence

Digital TV involves the production, distribution, and broadcast of programming in a digitally encoded format. According to the Federal Communications Commission (FCC), digital cable TV systems refer to those containing one or more channels utilizing the Quadrature Amplitude Modulation (QAM) modulation technique for transporting programs and services from a headend to a receiving device. **A digital cable TV network** is defined as networks using fiber optic cables to deliver digital TV, telecommunications, and Internet access. At the subscriber's end, it provides a unified platform and information gateway on TV screen with an interface similar to PC. In 2007, there were 36,200,000 digital cable households, accounting for 58% of basic cable subscribers and 32% of total U.S. TV households (NCTA, 2007). The advent of digital cable TV networks reflects current trends in communication technologies that are shifting from pc-centric to network-centric, from access or infrastructure provision to content and value-added services, from one-way broadcasting to two-way interactive communication, and from channel-carrying to unified platforms.

The key concept that explains the current transformation of cable TV into a triple-play service network is convergence. Scholars (Atkin, 2002; Baldwin, McVoy, & Steinfield, 1996) defined convergence as the integration of digital audio, video, text, and data into unified media platforms. From a business perspective, media convergence means that the boundary among broadcast, cable, telephony, and the Internet is crossed. Accordingly, Pavlik (1999, p. 54) has defined convergent media as "digital media emerging from the convergence of computing (the interface/platform), telecommunications (delivery system), and traditional media (content)."

Some scholars (Baldwin et al., 1996; Fidler, 1997) argue that the significance of media convergence is the integration of communication technologies with network platforms. That is, convergence makes it possible for one network, such as the digital cable TV system, to provide multiservice to subscribers. This is the impact of convergence cable TV networks. As Bates and Chambers (2004, p. 191) put it, cable TV networks have evolved from a single-platform for relaying broadcast TV into a "multiplatform, advanced telecommunications provider" with services ranging from basic cable channels to high-speed Internet connection. Thus, digital cable TV networks represent a unified platform for delivering a large number of TV channels, add-on services, telephone services and the Internet over fiber optic cables.

The major technological, political, and economic influences that facilitate the convergence of digital cable TV networks with telecommunications networks and the Internet are: digitization and compression technologies, government policies and deregulation, and service bundling as a new business model.

Traditionally, publishing, broadcasting, telecommunications, and the Internet were different communication technologies with each operating on its unique platform (i.e., newspapers and magazines, TV and radio sets, the plain old telephone, and PCs). The conversion from analog to digital form of publishing, broadcasting, telecommunicating, and data processing paves the way for media convergence (see Section 2.2). More important, digital broadcast technologies support a variety of screen formats (e.g., TV, PCs, PDAs, mobile phones, and MP4 players). They are flexible in program delivery and offer interactive applications. These attributes led to the development of a unified media platform. Bates and Chambers (2004, p. 180) characterized digitalization as the "third incarnation" of the cable TV industry.

Deregulation is another key factor that made convergence a reality (Fidler, 1997). Cable TV networks were defined originally as CATV under the supervision of state and local governments. The *1984 Cable ACT* made FCC and Congress the national bodies to regulate cable. To prevent cross-ownership, broadcast and telecommunications were regulated as separate industries and their services were distinct. Cable companies were defined as cable TV operators; while the telecommunications companies were known as common carriers. However, the *1996 Telecommunications Act* abandoned the decades-old policy. The Act was described as a milestone in the history of telecommunications (**Economides, 1999**) because it radically restructured the U.S. telecommunications markets by promoting new competition and greater choice for consumers. Telephone companies can now offer broadcast services; cable companies can offer telecommunications services.

As a result of deregulation, "the services that are now provided by separate companies—cable TV, telephone, broadcasting, and so on–are offered in tandem by each" (Reagan, 2002, p. 65). By unifying the services of cable TV, telecommunications, and the Internet into one platform, the convergent digital cable TV networks have become greater than the sum of three individual networks. A new business model building on the notion of synergy has emerged since 2000 (Wirth, 2006). It bundles cable, telephone, the Internet access in a single package. This new business model in providing 3-in-1 services, in turn, accelerated media convergence. The era of triple-play has arrived.

As Chan-Olmsted (1998) found in a study of mergers and acquisitions in the cable and telephony industries as a strategy for convergence in post-*1996 Communications Act* years, the American telephone industry was the most active in mergers and acquisitions. The U.S. cable industry was found to purse a horizontal consolidation strategy through internal mergers and acquisitions so as to be prepared for competition from an integrated telecommunications industry.

Digital Cable TV Networks: Architecture, Technologies and Services

Because of technological upgrades and media convergence, digital cable TV networks are now capable of offering a wide range of expanded functions in a single network. Compared to traditional cable TV networks that basically charge subscribers for receiving TV programming via wired cables (Bates & Chambers, 2004), digital cable TV networks are more than a distribution network. They are a unified platform and information gateway. The architecture, technologies, and applications of digital cable TV are discussed next.

The architecture of digital cable TV networks with a unified platform consists of two components: (1) hardware and (2) software. The hardware of digital cable TV networks includes a cable headend that receives digital programming and services from satellites or microwave links and then transmits them to subscribers via fiber optic cables. A subscriber control system that is placed outside the subscriber's home controls access to cable signals. In the home, it consists of a digital set-top box (STB), a cable modems for high-speed broadband Internet connection, and a digital TV set. New digital TV sets that have an expansion port can use a cable card to receive digital TV without a digital STB.

The digital STB is the most important piece of hardware on the subscriber's end because it executes critical functions. According to Pagani (2003, pp. 69-70), as a receiver, the digital STB is a subsystem with three general functions: (1) conditional access. That is, STB controls access through scrambling and encryption to guard against unauthorized use; STB also handles signal conversion from analog to digital; in addition, it transmits HDTV signals through a tuner; (2) it services as a return or interactive channel. That is, as a point-to-point link, STB allows two-way communications between the subscriber and the cable service center; premium and fee-based services such as VOD, NVOD, PPV, and DVR, go though it. So do interactive features of on-screen TV Guide, parental control, program alerts, interactive TV voting, gaming, and shopping; and (3) it is an electronic program guide (EPG). It is clear that the platform of digital cable networks relies on STB as its brain to function.

Industry research on the shipment of digital terrestrial TV (DTT)–enabled STB varies, fluctuating between seven and ten million from 2001 and 2003. Shipments worldwide increased sharply to 18 million in 2004. With massive digital upgrades of cable networks over the world in the past few years, especially the approaching deadline of digital conversion in the United States on February 17, 2009, demand will increase. A total 200 million digital STBs is estimated to be shipped worldwide by 2011.

The software of digital cable TV networks involves SMS (subscriber manager system), middleware, and an interface known as Multimedia Home Platform (MHP) to receive and execute two-way applications on a digital TV. Though digital cable networks deliver sharper pictures, enhanced sound quality, and a huge number of TV channels, the interface of digital cable TV networks resembles more PC than TV. The navigating device of digital cable TV services is EPG—Electronic Program Guide.

EPG is navigational device that provides program schedules. It also allows users to search programming by category, showing time or theme. By navigating through an EPG on a TV screen or the digital STB, subscribers can view information about a program being shown and forthcoming programs (call it the digital remote control with an on-screen display). EPG has a graphic user interface, which shows program titles, channel, and schedules on a grid. Subscribers have the option to select more information on each program, which can be VOD, PPV, NVOD or HDTV. Then, they can send the request to the cable service center. Also, they can activate DVR to record a program.

VOD and PPV are interactive technologies that allow mass customization of cable programs and content, while DVR facilitates time-shifting in viewing recoded programs. Figure 1 shows the configuration of the basic elements of the digital cable TV architecture.

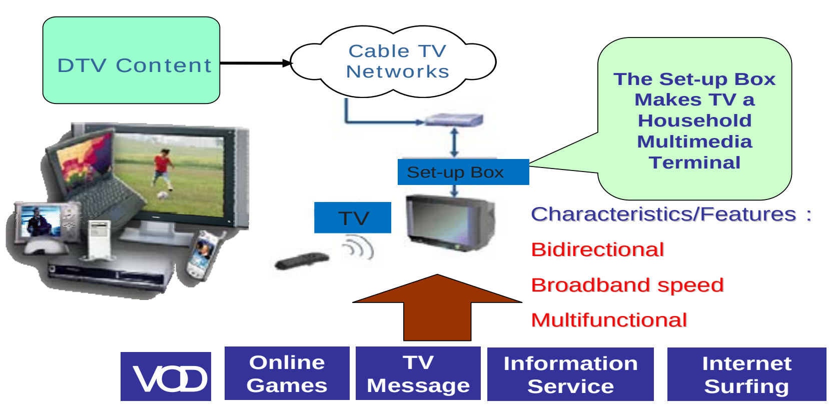
VOD refers to an interactive broadcast technology that enables subscribers to select videos from a central server for viewing on TV or PC. The scope of VOD applications on digital cable TV networks is broad. It can be used for ordering movies, TV shows, sports games, etc. It can also be used for requesting special programs or videos devoted to education or training. VOD turns out to be the most popular interactive video service all over the world, giving rise to an on-demand model of marketing and advertising.

PPV refers to an interactive service on digital cable TV networks. It enables subscribers to purchase premium shows to be viewed on a terminal like TV or PC. On digital cable TV networks, shows or special events can be purchased via an on-screen guide and EPG.

NVOD is an interactive TV technology that combines VOD and PPV. It enables viewers to see a PPV offering at staggered times (say every 15 minutes). Basically, it increases the opportunities for the same show or movie to be viewed.

HDTV (High-definition television) represents a new and higher TV broadcast standard with higher quality video, audio, and a wider image aspect ratio than standard TV broadcast signals. Digital cable TV networks have come onboard the HDTV bandwagon with the capacity of delivering HDTV signals on its overhauled infrastructure. Through an HDTV-ready STB with a built-in QAM tuner, digital cable subscribers can receive HDTV channels from their cable companies (usu-

Figure 1. The architecture and technologies of digital cable TV networks



ally designated as "cable TV HDTV channels"). In 2006, nearly 96 million U.S. households were passed by a cable operator that offers HDTV programming (Patterson, 2007). There is more HDTV programming available over digital cable TV networks than any other network.

Digital STB with digital recording function allows subscribers to use DVR, which captures digital programming and playback on digital cable TV networks (newer models of digital TV sets have built-in DVR hardware and software). With DVR services, TV viewing is never the same because DVR is constantly recording the channel that a subscriber tunes to. He or she can go back to the beginning of the program, never missing anything. Also, the subscriber can save the recording and view it another time, a time-shifting function similar to VCR (Video Cassette Recorder). The difference is that DVR can record several programs simultaneously, while VCR cannot. Finally, digitally recorded programs can be transferred to PCs and other devices such as PDAs or MP4 players for shared viewing.

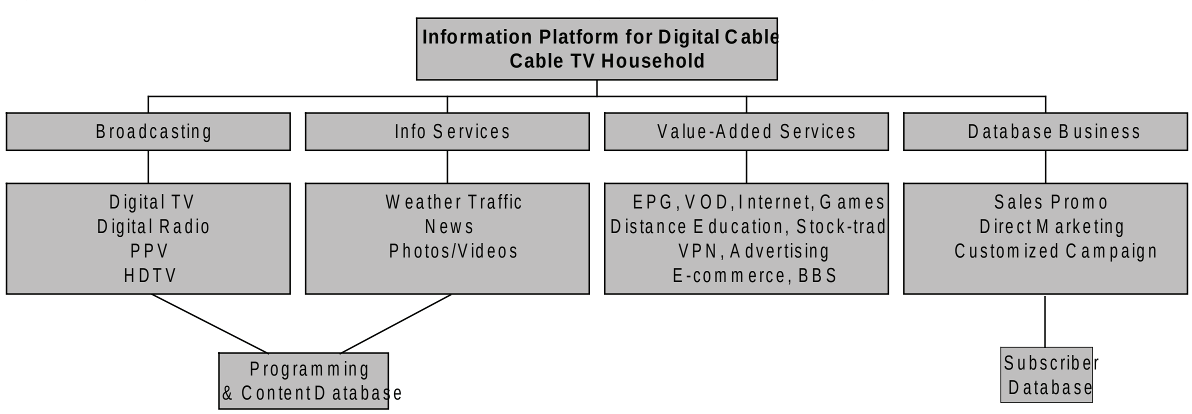
In sum, digital cable TV networks offer expanded channels, more choices through on-demand offerings, customized premium options, personalized interface, and always-on Internet connectivity. All of these services are offered over a single network. In other words, the digital cable TV network is nothing like the traditional one-way, analog cable TV network restricted to distributing TV signals. It is not a surprise that the current transformation of cable TV networks is described as a digital cable TV revolution.

As illustrated in Figure 2, as an information gateway, digital cable TV networks can provide such information services as digital teletext, news tickers, stock tickers, and live traffic updates. In addition, local listing of businesses and directory services can be provided to subscribers. Two-way video conference is another example. As a unified platform, the digital cable TV network can offer telephone and Internet-based services as value-added services, including classified, community forums, TV e-mail, multimedia interactive games, digital radio channels, e-commerce, and sites for user generated content (UGC) to share photos and video.

EMERGING BUSINESS MODELS IN MANAGING DIGITAL CABLE TV NETWORKS

Media convergence and government deregulation of the broadcast and telecommunications industries have created a favorable business environment for developing new business models. With a unified media platform that combines TV and PC, a wide range of value-added services and applica-

Figure 2. Digital Cable TV Networks as Media Platform and Household Information Gateway



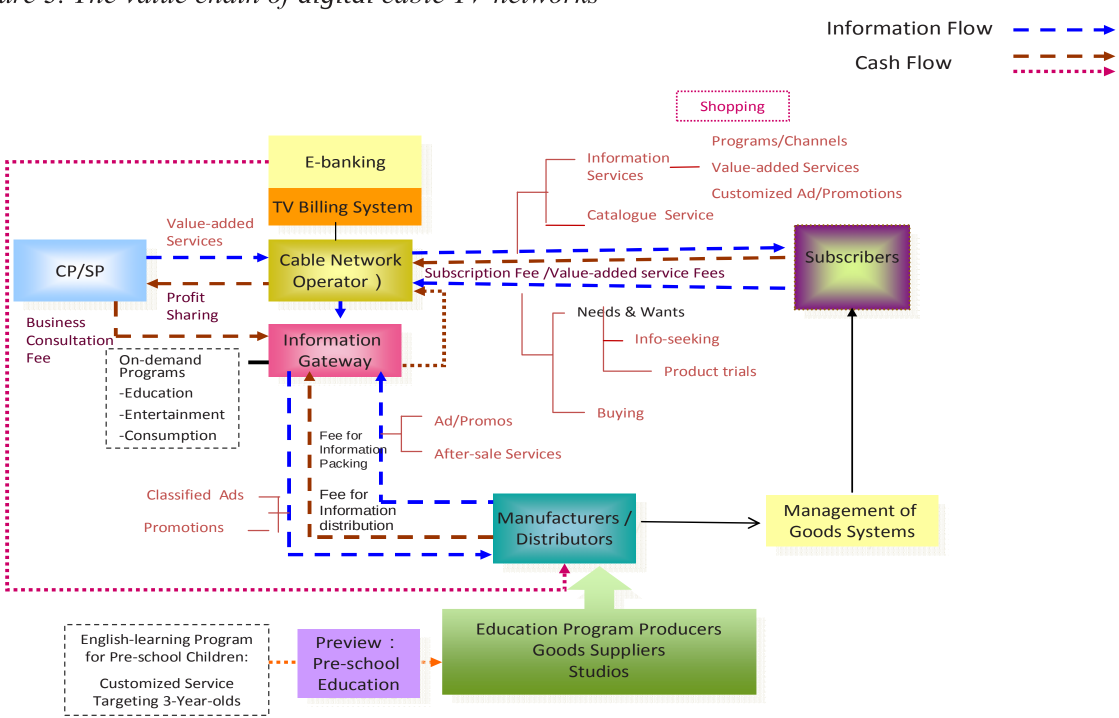
tions on digital cable TV networks has emerged. This section identifies key players in the digital cable TV industry, examines the value chain of digital cable TV, and proposes new approaches to manage the digital cable TV networks.

Players in the Digital Cable TV Business

As a convergent network of broadcast, telecommunications, Internet broadband connection, and Web applications, the business of digital cable TV is much more complicated and challenging than a cable system solely for signal transmission. The stake holders in digital cable TV include network operators, hardware suppliers, software vendors, and content providers and aggregators. Together, they create a value chain for the digital cable TV business (see Figure 3). It is worth noting that the digital cable TV value chain integrates with the Internet value chain. The collaboration among the participants in the digital cable TV business will foster the growth of new services in e-commerce, telecommunications, and Web applications such as IPTV.

Cable companies. They are known as MSOs—multiple-system operator—in the United States because cable companies own and operate numerous cable systems in different cities across the country. In 2007, there were 6,635 cable systems in the United States (NCTA, 2007). They are in the business of offering a variety of broadcast, Internet, and telecommunications services to American TV households for a charge. Paid services include cable subscription (basic and premium), HDTV cable channels, premium and interactive services (VOD, PPV, and DVR), broadband Internet connection (similar to an ISP—Internet Service Provider), and VOIP telephone services (similar to an ITSP—Internet Telephony Service Provide). As access providers, MSOs have the advantages of "point of presence" in reaching a subscriber's

Figure 3. The value chain of digital cable TV networks



home (Keen, 1997, p. 163); they charge cable content providers with a fee for delivering the point of presence.

Hardware suppliers. As discussed earlier, the digital STB with tuning capabilities is an essential device for the digital cable TV network. It processes broadcast signals on cable TV networks and feed them into digital TV sets. STB manufacturers sign agreements with MSOs to supply digital STBs. As digital conversion fastens its pace to meet the February 17, 2009 deadline, demand for digital STB will likely increase. However, technological standards do not exist in digital STBs. The technology of several popular models is proprietary. Therefore, digital STB is a highly competitive market. In the United States, the digital cable STB market is adopting a technology developed by CableLabs, an industry consortium of cable operators and equipment manufacturers through the OpenCable Initiative. Two manufacturers—General Instrument (acquired by Motorola in 2000) and Scientific-Atlanta—dominate the market. Other consumer electronics makers such as Sony, Philips, and Panasonic also make digital STBs.

Software vendors. Converging with the Internet, digital cable TV networks are capable of offering various Web applications, including email, chat-room, TV polling, video games, e-commerce (e.g., banking, e-bay, classified, job database, etc.), database services (e.g., phone directory, restaurant guides, etc.), and interactive education programs. These applications require special expertise and technological know-how. MSOs need to collaborate with vendors to make these services available to subscribers. For instance, systems to manage orders, billing, and delivering goods need to be outsourced to software providers.

Cable content providers and Web content aggregators. As early as the 1970s when the first cable-only channel HBO debuted, the cable content industry existed. They have grown significantly. According to NCTA (2007), the number of national video programming networks and services reached a record high of 565 in 2007. In 1995, the total was only 129. Content providers have strong competitive advantages in the era of media convergence and multiple media platforms. Content is always in high demand. The old saying of "content is the king" rings more true than any other time.

Moreover, as an information gateway, digital cable TV networks are empowered with Web publishing capabilities to distribute and store news and information over IP. However, MSOs are not in the news or publishing business. Therefore, to build a Web content business, they rely on content providers to aggregate timely and diverse news and information. The content may include traffic, weather, stock updates, and entertainment (e.g., movies, music, or games), among other things.

New Business Models of Managing Digital Cable TV Networks

As Keen (1997) points out, the competition in the traditional cable TV industry plays out among carrier/access providers, content providers, and total providers (both access and content). With media convergence, it is understandable that MSOs are concerned about competition from new players such as Webcasters, which may bypass cable TV networks to distribute video and TV programming on the Internet (Owen, 2001).

However, with an explosive growth of channels and media outlets, users probably play the most important role in affecting the business of digital cable TV. This is because media convergence empowers users, changing their role from passive viewers to active information seekers and value adders. As some scholars argued, broadcasting is shifting from "narrowcasting" to a "user as participant" model of communication (Slot, 2007, p. 305). In the context of using digital cable TV, it is no longer a layback but lean forward culture (i.e., users actively access media content). Stipp (1999) also suggested that consumer preferences will be "decisive" (p. 11). In other words, users are in charge.

Converged networks, new services, new platforms, and the new role of users call for new approaches to manage digital cable TV networks. The following are four user-centered models in managing the digital cable TV business to seek new revenue streams and increase profitability:

On-demand model. Industry research forecasts that most purchases of video content will be driven by consumer-demand applications by 2016. This model is based on users' abilities to select and choose programming and services over the unified digital cable TV platform. As digital cable TV networks have the competitive advantage in offering added on services such as VOD, HDTV, and DVR, the potential of new revenue streams out of these new services will be great. Specifically, in addition to offer basic and premium channels through VOD, NVOD, and PPV, MSOs can provide more options in value-added content such as educational materials and Web contents. Taking advantage of the interactivity of digital cable TV systems, this model seeks to create multiple revenue streams. Take EPG as an example. It is an interface similar to a Web portal on TV screen where users start out, interactively use cable TV, and turn off the system. Thus, EPG generates the most traffic, which can be sold to advertisers. STB, a unique two-way communication technology of digital cable TV, may also be used to generate new revenues out of placing digital ads integrated with the system.

Bundled service model. With the "one-stop" service concept, this model bundles broadcast (radio and TV, including HDTV), telecommunications (including wireless and OVIP), and Internet broadband services in one package. MSOs have enjoyed the growth of high-speed Internet connection using cable modems. In fact, the cable modem business was a cash cow for them. The growing business of providing local telephone services and VOIP added new revenues to recover some of the high costs in network upgrades. Moreover, MSOs have the opportunity to provide value-added services such as e-commerce transactions (paying bills, tickets, hotels, etc.). Revenues will come from cable subscription fees, add-on services, premium content, and advertising. This model is the prevailing model in the United States, and it is increasingly found in other countries. Its advantage is that it will likely increase the overall value of a digital cable TV system.

Mass customization model. Research (Jeffres, Kimberly, Neuendorf, & Lin, 2004) on user content choice behavior in today's converged media environment shows that their diverse interests influenced the actual media use and choice of content. This model focuses on cultivating major market segments and provides programming and services targeting each segment. Individual users of a certain segment are offered the option to custom-order programs and services to meet their specific preferences and interests. The basic idea is to allow users to build personalized media platform on which they can view TV, make phone calls, access the Internet, and publish online over digital cable TV networks. MSOs have competitive edge in targeting households and blocks, but they have not taken advantage of subscriber base to target regional markets. Doing so will broaden the appeal of cable as a regional medium to advertisers.

Community-building model. Digital cable TV networks are typically local. Thus, they are in a unique position to be the network for local communities in the capacity as an information gateway or Web portal on TV screens. Traditional cable TV networks are one-way and closed systems. But digital cable TV networks are open and interactive. MSOs can seek business opportunities in building an open environment on their systems to serve local communities. For example, they can create channels for directories, facilitate citizen journalism (e.g., loading pictures, posting comments, and social networking), and build links to corporate Web sites. Good corporate behavior is good business. Buzz advertising can be a new revenue stream under this model.

THE FUTURE OF DIGITAL CABLE TV NETWORK MANAGEMENT

The transformation of cable TV networks driven by digital communication technologies will continue in the years ahead. With digitalization, cable TV networks, which have reached the stage of maturity in its life cycle (Picard, 2002), are being reinvented. There is nothing like it in the nearly 60-year history of cable TV. On the other hand, MSOs face a competitive disadvantage in the transition to digital cable TV networks as a triple-play service provider. Unlike the VIOP over the Internet, MSOs need to overhaul the existing cable TV infrastructure, which cost millions to build and maintain. Thus, it should be noted that the capital required for system upgrades will have a major economic impact on MSOs.

In addition, the digital cable TV business has become complex, sometimes chaotic and uncertain. For example, there is a trend toward cable TV value chain fragmentation as more and more players participate in the digital cable TV business. Challenging issues in managing the digital cable TV networks include strategic planning (e.g., how to position digital cable TV networks), technological upgrade (e.g., wireless cable), packaging and marketing of value-added services to target markets, assessing market demands, and findings innovative ways to serve customers and communities.

New tools for managing digital cable TV networks are also emerging. For example, data mining and database marketing. A wealth of data can be collected in digital cable TV networks, which can be turned into information to identify marketing and advertising opportunities. Database marketing is another management tool that helps reach key segments with customized service bundles targeting them.

REFERENCES

Atkin, D. (2002). Convergence across media. In C. Lin & D. Atkin (Eds.), *Communication technology and society* (pp. 23-42). Cresskill, NJ: Hampton Press.

Baldwin, T., McVoy, D., & Steinfield, C. (1996). *Convergence: Integrating media, information and communication*. Thousand Oaks, CA: Sage.

Bates, B., & Chambers, T. (2004). The cable industry. In A. Alison, J. Owers, R. Carveth, C. Hollifield, & A. Greco (Eds.), *Media economics: Theory and practice* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum & Associates.

Chan-Olmsted, S. (1998). Mergers, acquisitions, and convergence: The strategic alliances of broadcasting, cable television, and telephone services. *The Journal of Media Economics, 11*(3), 33-46.

Cicora, W., Framer, J., & Large, D. (1999). *Modern cable television technology: Video, voice, and data communication*. San Francisco, CA: Morgan Kaufmann.

Collette, L. (2004). Cable telecommunications. In A. Grant and J. Meadows (Eds.), *Communication technology update*. NY: Focal Press.

Damascio, M., & Ferreira, A. (2004). Interactive television usage and application: The Portuguese case study. *Computers & Graphics, 28*, 139-148.

Economides, N. (1999). The Telecommunications Act of 1996 and its impact. *Japan and the World Economy, 11*, 455-483.

Fidler, R. (1997). *Mediamorphosis: Understanding new media*. Thousand Oaks, CA: Pine Forge.

Guillory, B. (2006). Cable television. In A. Grant and J. Meadows (Eds.), *Communication technology update* (pp. 70-82). Boston, MA: Focal Press.

Harte, L. (2007). *Introduction to cable television: Analog and digital television and modems.* Fuquay Varina, NC: Althos Publishing.

Jeffres, L., Atkin, D., Neuendorf, K., & Lin, C. (2004). The influence of expanding media menus on audience content selection. *Telematics & Informatics*, *21*, 317-334.

Keen, P. (1997). *Business multimedia explained: A manager's guide to key terms & concepts*. Cambridge, MA: Harvard Business School Press.

NCTA (National Cable & Telecommunications Association). (2007). *Statistic*s. Washington, DC: NCTA. Retrieved January 15, 2008, from http://ncta.com/ContentView.aspx?contendID=54.

Nielsen Media Research. (2006). *Cable TV audience: 2006 annual report*. Retrieved on January 22, 2008 from http://www.journalism.org/node/509

Owen, B. (2001). *The Internet challenge to television*. Cambridge, MA: Harvard University Press.

Pagani, M. (2003). *Multimedia and interactive digital TV: Managing the opportunities created by digital convergence*. Hershey, PA: Idea Group Publishing.

Patterson, B. (August 16, 2007). Three ways to get HDTV programming. *CNET HDTV World*. Retrieved January 24, 2008, from http://www.cnet.com/4520-7874_1-5108854-2.html

Pavlik, J. (1998). *New media technology: Cultural and commercial perspectives.* Needham Heights, MA: Allyn & Bacon.

Picard, R. (2002). *The economics and financing of media companies.* Bronx, NY: Fordham University Press.

Reagan, J. (2002). The difficult world of predicting telecommunication innovations: Factors affecting adoption. In C. Lin & D. Atkin (Eds.), *Communication technology and society* (pp. 65-90). Cresskill, NJ: Hampton Press.

Slot, J. (2007). Changing user roles in ICT developments: The case of digital television. *Telematics & Infomatics, 24*, 303-314.

Stipp, H. (1999). Convergence now? *The International Journal on Media Management, 1*(1), 10-13.

Symes, P. (1998). *Video compression: Fundamental compression techniques and an overview of the JPEG and MPEG compression systems.* NY: McGraw-Hill.

Wirth, M. (2006). Issues in media convergence. In A. Albarran, S. Chan-Olmsted, & M. Wirth (Eds.), *Handbook of media management and economics* (pp. 445-462). Mahwah, NJ: Lawrence Erlbaum & Associates.

KEY TERMS

Cable TV Networks: A system of transmitting and distributing broadcast signals to TV households using a network of fixed cables.

Convergence: Digital audio, video, text, and data are integrated into unified media platforms.

Compression: Reduction of the quantity of digital data used to represent video images.

Digital Cable TV Networks: A cable TV distribution network using digitally compressed video over fiber optic cables.

Digital Format: Information presented as a series of zeroes and ones.

DVR (Digital Video Recorder): A video recording technology that captures digital programming and playback on digital cable TV networks.

EPG (Electronic Program Guide): A navigational device that provides program schedules, which allows subscribers to search programming by category, showing time or theme.

HDTV (High-definition Television): A new and higher TV broadcast standard with higher quality video, audio, and a wider image aspect ratio than standard TV broadcast signals.

Near Video On Demand (NVOD): A technology that enables viewers to view a PPV offering at staggered times.

PPV (Pay Per View): An interactive service on digital cable TV networks, which allows subscribers to purchase premium shows to be seen on TV or PC.

STB (Set-Top Box): A digital device that controls cable access through scrambling and encryption; it also handles signal conversion from analog to digital and performs other key functions of interactive TV.

VOD (Video-On-Demand): An interactive broadcast technology that enables subscribers to select videos from a central server for viewing on TV or PC.

Chapter XXXVI
Diffusion and Oscillation of Telecommunications Services:
The Case of Web 2.0 Platforms

Tobias Kollmann
University of Duisburg-Essen, Campus Essen, Germany

Christoph Stöckmann
University of Duisburg-Essen, Campus Essen, Germany

Carsten Schröer
University of Duisburg-Essen, Campus Essen, Germany

ABSTRACT

The diffusion of a Web 2.0 product or services is, unlike to traditional consumer or industrial goods, not only based on purchase. Full acceptance of Web 2.0 platforms occurs by recurring utilization. The chapter focuses on diffusion characteristics of this innovative category of ICT products and provides management concepts for competition. The concept of critical mass is applied to different growth scenarios. Additional success factors are discussed. Particularly the permanent supervision of a platform regarding its compliance with qualitative, as well as ethical and legal standards is of great importance. Adjustments to external market conditions, proactive management, and a bilateral marketing approach are a key to lasting success within the Net Economy. Markets are never settled, due to the ever changing and oscillating conditions. The chapter shows that there is always a chance to capture a market or at least to grow against competition in a Web 2.0 setting.

CRITICAL MASS AS A SUCCESS FACTOR

The extension of electronic networks and the use of information and telecommunication technologies for the digitalization of value creation lead to a new economic dimension (Lumpkin & Dess, 2004). This newly established level of value creation, the so-called **Net Economy**, provides room for innovative business models and successful start-up firms (Kollmann, 2006). An increasing number of companies participate in the economic potential of the internet which leads to a rising level of competition. Competing players either win a market and participate in a stable and sustainable business development or fail with their idea within a short period of time (Shapiro & Varian, 1999). The roots of this phenomenon are derived from an economies of scale effect that keeps aggravating itself instead of declining. Every new user of an offered **platform** (community or marketplace) helps to raise the value of a network and makes it even more attractive for further participants. A higher number of communication and transaction activities are the possible outcome. A rising quantity of community members also increases the perceived attractiveness (site stickiness) of a platform. This can be illustrated by the following two examples: A rising number of members subscribing to an ***E-Community*** (Kollmann, 2006) raises the chance to meet likeminded individuals or to receive answers to posted questions. Also a rising number of users to an ***E-Marketplace*** (Kollmann, 2006) rises the probability to find interested customers for offered products of a supplier.

According to the presented scenarios, a special focus has to be put on the **critical mass** phenomenon, because the subjectively perceived attractiveness of a system (e.g. community) is highly correlated with the already registered number of users. A certain number of users within a network are necessary to create value among the participants at a sophisticated level. Reaching this level is essential for a network, because the enrolled participants will be reinforced to use the system on an ongoing basis, and it will become easier to convince new users to join in (Kollmann, 1998). The minimum number of participants to maintain a sufficient utility on a long-term basis is referred to as the **critical mass** (Weiber, 1992).

Especially in a **Net Economy** setting young companies experience a very competitive environment to reach the **critical mass** (Kollmann, 1998). Oftentimes, the winners of this race drive smaller competitors or copycats off the market. This conception reinforces itself in a **Web 2.0** setting (O'Reilly, 2005), where customers or members leave the status of pure information consumers. Their status changes to an active information provider and editor role (O'Reilly, 2005). Therefore, growth at a fast pace in regards to the number of users becomes the critical success factor to leave the zone of competition as a winner. Actually, the winner of this battle is able to establish a close too monopolistic market position (Shapiro & Varian, 1999). The attractiveness for new users to join a network is even higher, if *everyone else* already joined in.

Following the stated assumptions Web 2.0 **critical mass** winners are destined for lasting company performance and profits. But the real life teaches another lesson. Apparently successful market leaders are frequently challenged by various inconveniences with the potential to jeopardize their market position. In accordance with the theoretical model ***eBay™*** for example, market leader for internet auctions, announced a growing number of membership accounts alongside with rising revenues and profits (eBay™, 2006). The unmentioned downside of this success story was a flood of insolvencies among professional ***eBay™*** dealers. *The International E-Business Association* (IEBA), an association for *power sellers*, sees the roots for many discontinued businesses closely connected with an increasing number of sellers and a resulting higher level of competition. Both factors lower profits and force sellers to predatory pricing strategies. In this context insolvency reasons for most of the dealers are not on an individual, entrepreneurial level. They are based on the market characteristics of electronic marketplaces, and a substantial number of insolvencies by professional dealers of a platform will sooner or later hit the marketplace vendors.

Other **critical mass** winners within the **Web 2.0** environment like the online community ***MySpace™*** or the video **platform** ***YouTube™*** are not only the centre of interest because of their enormous growth rates and success stories. Critical notes about security issues, copyright violations, or identity theft and fraud affairs are also on the spot of public interest. Web 2.0 companies might face severe challenges, if the offered content on their platforms violates ethical or legal standards. Insufficient qualities of the offered content, as well as a mismatch between information supply and demand in the case of ***eBay***^{TM}, gain the potential for adverse effects for a market position.

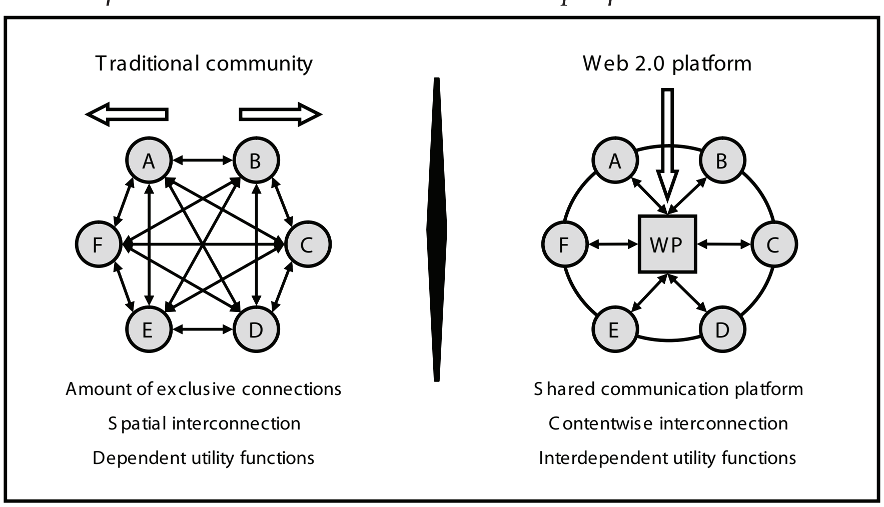
Therefore, the proposed chapter aims to show how quantitative, qualitative, as well as ethical and legal matters correspond with the market success of Web 2.0 platforms. In addition implications for the competition of platforms and the concept of **critical mass** as a foundation for success in the **Net Economy** will be discussed.

WEB 2.0 PLATFORMS IN THE NET ECONOMY

In the past the internet used to be recognized as a technology to publish and distribute data, information and media content. This view was based on split-up roles: Private and commercial publishers of web content with an active role on the one hand, and passive consumers on the other hand. This *golden rule* changed in 2005, when **Web 2.0** saw the light of day. A new category of website concepts was born. The established differentiation between active content providers and passive consumers diluted. Now users were able to generate and affect contents. User generated content became the slogan for the new internet. The active role of users built the basis for innovative business ideas, which were unthinkable some months ago. Many **Web 2.0** business models, like the online community ***MySpace***^{TM} or the video **platform *YouTube***^{TM}, are centered around community structures. According to Kollmann (2006) an **E-Community** facilitates contact and interaction between individuals or institutions via a digital network. Therefore, the integration of innovative ICT supports data- and knowledge transfer. These two features characterize the core activity for most of the **Web 2.0** business models. Besides arranging and exchanging contacts and information, bringing together supply and demand for economic transactions, e.g. on an **E-Marketplace** (Kollmann, 2006), is an integral feature of numerous platforms.

The common goal of **Web 2.0** platforms is to win suppliers and consumers of information for their business model in order to match them

Figure 1. Illustration of traditional communities and Web 2.0 platforms



(Kollmann, 2006). Users of a **Web 2.0** platform act alternating and parallel as information suppliers and consumers. Both activities have to be considered separately, because supplying and requesting information differ in regards to motivation and acceptance. This leads to a tripolar structure. The platform operator provides a matching service to perform an exchange of information or a business transaction at lower transaction costs (Lee & Clark, 1996). The value of a **platform** not only depends on the operators' service capability and willingness to perform, but also on the contributions of the suppliers and consumers of information (derivative capability aspect). Platforms depend on queries. A higher number of queries provide a broader scope for matching activities (Kollmann, 2001). Consequently a **Web 2.0** platform is solely dependent on the participation willingness (acceptance) of its users. Therefore business development efforts concentrate on the so-called matching as a target parameter (Kollmann, 2000, 2005).

When starting a **platform**, operators have to get awareness for their services. They are challenged by the question, which points are of interest to turn internet surfers into members or subscribers for a community, or respectively customers for a marketplace. This goal can only be reached by offering a matching **platform** that delivers an acceptable service. Therefore the scientific construct of acceptance obtained growing relevance in the marketing of ICT products and services over time (Kollmann, 2001). The reason for this is that ICT technologies and applications (e.g. interactive TV, internet, cellular phones) need a specific pattern of utilization. The pure purchase is not an indicator for further activities of a user and therefore not a sufficient indicator for the economic success of a **platform** operator. Augmenting this idea, a full acceptance of a **Web 2.0 platform** is closely linked to three conditions.

1. Connecting (Access to the platform): The customers have to get access to a platform via a security code (e.g. log-in) or provided access software. A first time registration process reflects the purchase.
2. Acting (Demand and contribution of information): The customers have to use the services of the electronic platform. Information has to be requested and provided. This opens up the potential for matching.
3. Interacting (Clearing and matching): The customers have to interact on the platform; otherwise a matching of requested and provided information cannot be executed.

Coordination of the participants is proceeded over the electronic **platform** (n information suppliers, m information consumers, and the platform operator without time and local limitations, see figure 1). Information and data between two or more counterparts is solely shared on the **platform**. The active placement and intervention into the matching process leads to a new responsibility for the platform operator, because the result of each matching affects all participants.

The attractiveness of traditional communities and marketplaces is primarily determined by the numbers of participants as a quantitative measure for their availability (see figure 1). The center of those system architectures is the exclusive connection between two counterparts (e.g. trade show or farmers market). The interaction between A and B usually has no direct qualitative effect for the utility of C. Technological external effects are conceivable under certain conditions, e.g. if all participants benefit from a network effect associated with the extension of a telephone network. However, this interpretation of a quantitative point of view is insufficient to explain attractiveness of progressive internet platforms.

In a **Web 2.0** system the interconnection of participants does not consist of exclusive one-on-one data links. An **E-Community** or **E-Marketplace** as a commonly shared platform represents the center of the system architecture (see figure 1). Information of a database is available for every user on the network. Options to alter, comment, or expand the provided content are inherent to the system. A transparent provision of information leads to direct effects on the qualitative utility function of C (economies of scope), if A and B exchange information. An agreeing or derogative comment of B to a contribution from A, may lead to a valuation of C regarding the statement. Further, as Pavlou and Gefen (2005) state, a psychological contract violation with an individual seller is proposed to prompt a generalized perception of contract violation with the entire community of sellers in a marketplace. Internet auctions for a specific item are open to several participants. The bid of A has a direct impact on the utility function of all other bidders. The following paragraph will deal with the resulting implications for the diffusion of Web 2.0 platforms.

THE DIFFUSION OF WEB 2.0 PLATFORMS

Research on **diffusion** provides answers to the question how an innovation will spread on a market (Rogers, 2003, Pavlou & Fygenson, 2006). Services of a **Web 2.0** platform, which are offered as a commercial product, could be part of a study on diffusion, too. Research in this field is based on the presumption of a recurring use (acceptance) of a product or service, not on a one-time purchase (adoption). In regards to the diffusion of **Web 2.0** platforms the following three questions are of interest.

1. Which factors have an impact on the diffusion of a Web 2.0 platform on a market?
2. How fast does a Web 2.0 platform spread on a market?
3. What are the growth characteristics of a network?

The successful **diffusion** of a **Web 2.0 platform** is completed, if all interactions of a defined market are handled by this platform. For the evaluation of **diffusion** the aforementioned quantitative alignment of the network effect (higher number of participants = higher probability to find appropriate counterparts for interaction) appears with an economies of scope effect (nature, size and trend of an executed transaction, including its impact on the overall system). The following paragraphs will analyze the main problem areas of diffusion and oscillating degrees of utilization in the light of both effects.

Problem Areas of Diffusion

The **diffusion** of a **Web 2.0** platform is associated with quantitative, qualitative, ethical and legal challenges. Those issues will be discussed in the following.

Quantitative Problem Areas

Attractiveness of a web **platform** is significantly linked to the number of participants. A higher participation level raises the chance to reach other individuals. Every information supply (e.g. a provided video) as well as every information demand (e.g. on a personal level) need at least one counterpart to enable a platform provider to match requests. The service of the platform provider creates an indirect utility that is derived from the usage of an interactive relationship within the communication system, the so-called derivative capability aspect (Katz & Shapiro, 1985; Farrell & Saloner, 1985). The derivative utility following the usage of such a good increases with the number of participants, and the intensity of use by the other participants (Weiber, 1992). The result is a network effect. Common examples for goods with direct network effects are all types of ICT-systems. The utility of each participant is advanced with every new customer, who helps to grow the network. In connection with the bilateral customer orientation (information supply and demand) of the platform provider, specific characteristics within the **dif-**

fusion of a **platform** can be derived for different development stages (Kollmann, 2001).

1. Chicken-and-Egg-Problem: One reason for matching problems on a **Web 2.0** platform is derived from the so-called Chicken-and-Egg-Problem (Durand, 1983; Earston, 1980). The following two examples aim to illustrate this circumstance. An insufficient number of suppliers or offers lead to an absence of customers on the platform. An insufficient number of customers or requests lead to a lack of suppliers. The dilemma situation, which counterpart (supplier or consumer) at first has to get involved with the platform, is deemed as an obstacle for the development of a business.
2. Collateral-Critical-Mass-Problem: The installed basis, i.e. the number of users already present in a platform, determines the utility of the platform for new users since a greater number of users also increases the number of potential interactions (Farrell & Saloner, 1986). The larger the installed basis, the larger is the derivative utility for the participants (Kollmann, 2001). **Web 2.0**-platform providers are confronted with a collateral **critical mass**, because of the bilateral orientation (Kollmann, 1998). Suppliers need a certain level of counterparts or requests, in order to commit to or use a marketplace. Simultaneously, a certain level of suppliers or offers has to be provided, in order to persuade a customer to facilitate a marketplace. This problem supersedes itself, if the customer base on both sides grows to a sufficient point, where the derivative utility exceeds a certain level.
3. Equilibrium problem: Bilateral matching results in a mutual state of dependence regarding the number of suppliers and consumers, and respectively their offers and requests. Consequently the platform provider has to take into consideration that offers and requests almost equate themselves. Bilateral marketing activities support this endeavor (Kollmann, 1998) and help maintaining a high matching level (one offer = one request).

Qualitative Problem Areas

Contrary to the established belief in an exclusive utilization act between supplier and consumer interaction, the critical phase of a matching includes an additional economies of scope effect associated with the quality of interaction. The decision to subscribe to and use a **platform** has to be expanded. Besides complying with quantitative issues, meeting the qualitative requirements of the suppliers and consumers with information is of equal importance. If they realize that the web platform complies with their demand and interaction needs, they are willing to utilize the platform's services. The following issues with regard to qualitative problems have to be solved (Kollmann, 2001).

1. Matching performance problem: An exclusive focus on the number of suppliers and consumers is insufficient to measure the quality of the interacting counterparts, as well as their level of satisfaction with regard to the exchange of information. The demanded level of interaction has to correspond with the expectations of the participants. The degree of satisfaction is closely linked to three core areas of need, which are information, relationships, and business (Hagel & Amstrong, 1997), as well as the related concept of the heterogeneity-dependent level of commitment. Participants look for like-minded individuals on a platform and relevant content to fulfill their information needs. New discussion threads have to be established to acquire additional members, which cover further fields of interest. Platform operators are exposed to a dilemma

situation, because of the diametrical impact of the heterogeneity of discussion threads with regard to a growing member base and the persistent commitment of current participants.

2. Reality check problem: The structural conditions of virtual platforms disallow to validate provided information with reality. Anonymous publishing options among some platforms aggravate this issue. Information and reality fall apart frequently. Some of those discrepancies occur inadvertently, e.g. if a change of address is not entered into a database or an information is provided on a non-current standard of knowledge. More frequently intentional misrepresenting takes place by sugar-coating one's profile on the web, or even worse with criminal intention. Because the roots of a reality gap are unknown to the participants, misrepresentations might reduce the commitment or ongoing patronage of a user.

Ethical and Legal Issues

The addressed willful misrepresentation is an example for the multitude of ethical and legal issues associated with user generated content, among other legal problems of internet **platform**s and their foundation. Current lawsuits on those topics will give answers to important questions and provide future guidance on duties of **Web 2.0** platform operators. Liabilities for provided content on a company's website and linked content from external sources, as well as infringements of users (e.g. announcement of a criminal offense) and related duties of care have to be clarified. The results of those decisions will inevitably have a major impact on the further diffusion and development of **Web 2.0** platforms. In the following section two problems will be discussed.

1. Freedom of expression problem: Despite the fact that every human being possesses the right to express an opinion, legal and ethical standards have to be obeyed. Sometimes those standards are violated on anonymous web-based communication platforms. Comments with an extremist, offending, or sexually harassing content cannot be tolerated by any platform operator. Other categories, like advertising and promotion activities, can be classified as unwanted too. Guidance for communication on the internet is given by the so-called *netiquette*, derived from internet and etiquette. The recommended behavior of the netiquette is not legally binding, but helps to maintain and develop a positive net culture. The voluntary agreed upon rules are frequently incorporated in codes of conduct of platform operators. A breach of the rules leads to a closing of discussion threads, cancelation of comments, or dismissal of accounts, because a negative communication culture bears the lasting potential to lower the acceptance of a platform.
2. Problems associated with the adoption of external content:

Particularly on video platforms users provide, intentionally or unintentionally, copyrighted material from other websites or real sources. By now, copyright holders mandate agencies (e.g. *copyrightcontrol*™) to retrieve their protected material. The platform operator is responsible for inflicting penalties of participants and to remove copyrighted materials. Preventing an upload of protected material is virtually impossible, because the violation of a copyright just becomes apparent after a user has posted illegal content. Even reactive behavior leaves a legal stain on the platform operator's vest, as despite the fact of a fast content removal a breach of law already happened. Since there is no appropriate method available on the market to avoid the upload of copyrighted material, all prominent platform operators strive towards general licensing agreements with bailees. Those arrangements would allow them to leave copyrighted material on their websites and protect them from costly lawsuits and negative impacts on their market penetration.

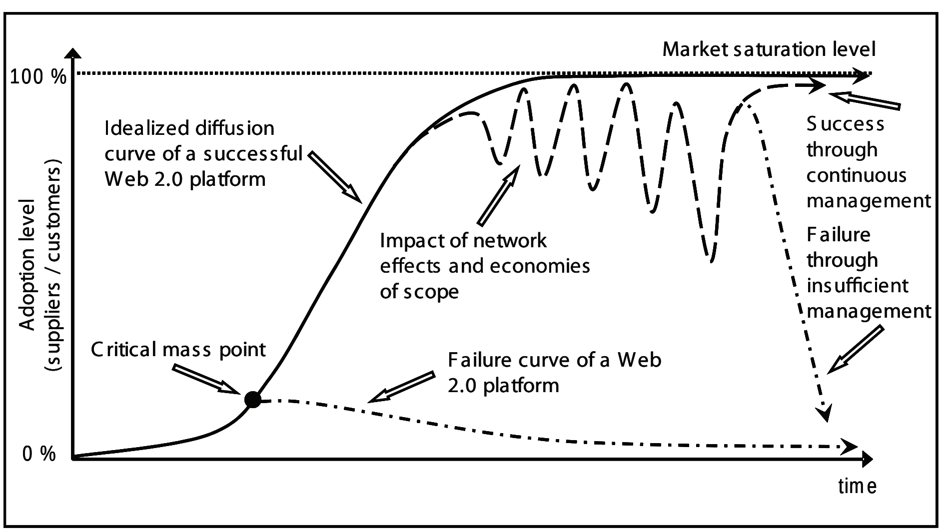
Oscillating Effects of Web 2.0 Platforms

The classical theory of **diffusion** was developed for so-called singular goods. Their diffusion proceeds with the act of buying. In a **critical mass** system the scope has to be extended by the variables of *connecting* and *acting* (see chapter *Web 2.0 Platforms in the Net Economy*) as constitutive determinates for diffusion (Weiber, 1992). With traditional consumer and producer goods the act of buying creates a positive and irreversible impact for diffusion. However, ICT service contracts bear the risk of being cancelled (e.g. mobile phone contract, website account), which limits the chance to realize demand synergies. As an extreme example reversible utility could cause a declining diffusion (see figure 2; Weiber, 1992). The characteristics of a **diffusion** curve in a **critical mass** system generally do not reflect a monotonically increasing function; in fact a considerable drop is also possible.

Diffusion Characteristics

The traditional model needs an expansion within a **Web 2.0** setting, because the connecting act is an insufficient parameter to evaluate **diffusion**. It is just a necessary requirement for adoption. The market success of a **Web 2.0 platform** depends directly on the participants' constant utilization and interaction as a reliable measure for adoption, and therefore acceptance (Kollmann, 2001). An adequate utilization and interaction discipline supports a premium quality of information and knowledge transfer among the participants of a platform with positive effects for the whole market system. Also the recurring utilization and interaction is a prerequisite to realize constant cash flows for the platform operators. Accordingly the concept of **diffusion** has to be extended beyond the purchase dependent quantitative measure of participants to the utilization and interaction dependent quantitative interaction measure. The reversibility of utilization and interaction has to be considered in this context. Due to the planning interval the sequence of the three adoption and

Figure 2. Diffusion of Web 2.0 platforms



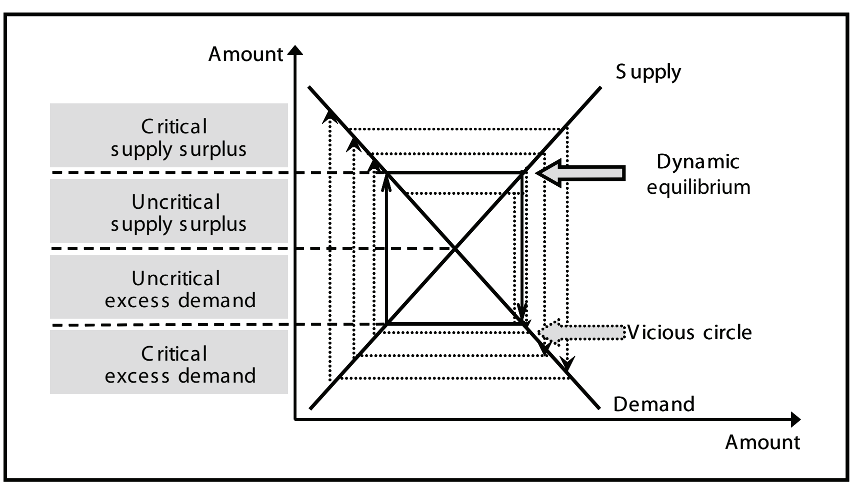
acceptance acts is interpreted as a discontinued multiple event. This process is characterized by permanent oscillations making **diffusion** a permanent companion. The reach of market saturation is not only linked with negative adoption and acceptance ratios, but also with alternating positive and negative ratios. The direct consequence is an oscillating development at market saturation level.

This effect is caused by the circumstance of negative utilization and interaction levels, which lead to a renouncing or deregistration by the participants. The original decision of adoption is withdrawn. Potential reasons to resign are caused by effects from quantitative, qualitative, ethical and legal issues. Suppliers and consumers of information feel uncomfortable about these aspects. The interconnection of participants has a negative impact on their individual utility function and the whole market system. An oscillating **diffusion** is the result (see Figure 2). In this context special attention has to be paid to the proportion of information supply and demand. As postulated, both factors should create an equilibrium to meet all matching requests and create a high level of satisfaction among the community members. The ever-changing level of participants leads to an alternating level of supply and demand. **Web 2.0 platform**s with a transaction oriented business model (e.g. ***eBay***^{TM}) use the price of goods as an instrument to regulate the proportions of supply and demand. Even though the control mechanism does not converge, due to continuously changing prices and numbers as well as suppliers and consumers, an indifferent dynamic equilibrium is the possible outcome (see figure 3).

Communication oriented platforms are faced with different challenges. A price based self-regulating mechanism does not exist in their business environment. Monitoring each market situation is important to perform intervening actions. From a platform operator's point of view the dynamic participation of suppliers and consumers is not necessarily associated with challenges. The dynamics reveal a harmful potential, if a critical supply surplus or excess demand is detected. Both scenarios cause the same effect (see figure 3).

The reduction of supply side participants will most likely lead to a decreasing information supply. Hence there is not enough potential to match all information requests. This causes a permanent loss of attractiveness on the demand side, as the offered information does not meet expectations. A lack of quantity and quality causes consumers to leave the **platform**. Likewise the lower number of requests makes it even more unattractive for

Figure 3. Adoption process of information supply and demand



suppliers to provide information. In this case the development of a platform does not induce an oscillation, but moreover a creeping and declining growth. Neither suppliers nor consumers of information are stimulated to return to the platform and utilize the services again (see figure 3).

In conclusion, the control mechanism of supply and demand generates a self-aggravating effect with a positive (virtuous circle) and negative (vicious circle) development potential (Kollmann, 1999). This control mechanism makes the management of a **Web 2.0 platform** complex, because the operator as an independent facilitator faces a bipolar user group simultaneously. Matching efforts should consider individual as well as general interaction requests to fulfil the expectation of opposed groups of interest.

The positive scenario (virtuous circle) in accordance with the **critical mass** effect leads to a continuous growth of power, achievement potential, and attractiveness of the platform. A significant gain of information suppliers typically shows an increase of requests with positive effects on the choice of the selected set of matching opportunities. In turn, the positive impact raises consumer satisfaction and the number of information requests. The flourishing demand for information lifts up the platform attractiveness for the suppliers and so forth (positive loop).

As stated before in this paragraph, changed market conditions may possibly restart a negative control mechanism. Within this vicious circle a significant loss of information suppliers leads to a clear reduction of requested information. Negative impacts on the number of choices for a matching lead to a high number of unmatched interaction demands. This has a negative impact on consumer satisfaction and consequently on the amount of information request. A declining demand reduces the attractiveness of the platform for suppliers, which leads to an ongoing downturn of supply and so forth (negative loop). The **Web 2.0** platform could suffer from those effects by a continuous loss of power and achievement potential. At the bottom line the existence of the platform might be at stake.

The oscillating characteristics of the **diffusion** curve lead to serious implications regarding the management of a **Web 2.0 platform** and the competition between web-based communication platforms in general. Those aspects along with the illustration of the **critical mass** effect as a success guarantee will be explained in the following paragraph.

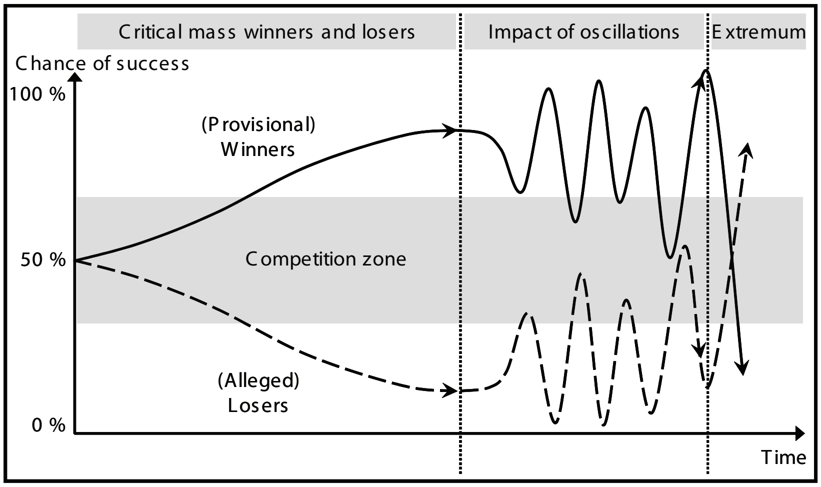
Competition on Diffusion

Competition in the **Net Economy** is characterized by an oscillating utilization of **platform**s. Therefore reaching a **critical mass** does not automatically assure a promising and stable development (see figure 4). Even on saturated or apparently settled markets threats for **critical mass** winners occur on a frequent basis. Alleged losers or innovative start-up companies occasionally create surprisingly good chances to grow against reputable competition.

As demonstrated, established **Web 2.0** platforms are endangered to lose participants. In a worst case scenario they start a vicious circle with the potential to bring their company down. The goal of successful **platform**s is therefore to maintain equilibrium between the bipolar groups as well as safeguarding compliance with qualitative, legal, and ethical standards. Lasting survival is closely connected with a high level of commitment among the participants and protection against competition. The continued management of a web **platform** becomes the critical factor of success. Provisional **critical mass** winners have to be on permanent alert, instead of relaxing in their accomplished position. Ongoing market evaluations and proactive influence on current developments on the respective network through bilateral marketing are inevitable.

Weak phases of critical mass winners provide opportunities for start-ups and established competitors to attack the supremacy of market

Figure 4. Competition in oscillating areas of conflict[1]



leaders. The existence of financial strength and survivability supports the gain of market shares (see figure 4). It is unrealistic to turn around a market completely or drive a leader in a certain field off the market, but addressing special target groups with innovative niche products receives growing popularity. Current examples are university and high school student communities, which recently started their services. Starting a positive control mechanism (positive loop) is the beginning to overcome the **critical mass** sustainably. Especially young start-ups are confronted with the challenge to reach **critical mass**. Their brand name is usually unknown to the broad public and the network attractiveness of their platform is limited. But the value of a network product is not solely based on the number of participants. Future development expectations also play an influential role (Hagel & Armstrong, 1997). Well timed and promising announcements to the market in advance (vapor marketing) and the management of expectations combined with additional online and offline marketing activities bear the potential to occupy a niche by the massive acquisition of new customers.

CONCLUSION

The **critical mass** plays a vital role for the implementation of internet business models. As described, the **critical mass** concept is of crucial importance in the age of **Web 2.0**. Current developments show further challenges besides this success factor. Those challenges have to be taken into account with the management of a **Web 2.0 platform**, because they could have a severe impact on company development. Particularly the permanent supervision of a platform regarding the compliance with qualitative, as well as ethical and legal standards is of great importance. Adjustments to external market conditions, proactive management of the platform, and a bilateral marketing approach are key for a lasting success within the **Net Economy**. Finally competitors and founders of (new) ventures should keep in mind that a market is never settled, because of the ever changing and oscillating market conditions. There is always a chance to capture a market or at least to grow against competition.

AUTHOR NOTE

Tobias Kollmann, Christoph Stöckmann, and Carsten Schröer, E-Business and E-Entrepreneurship Research Group, University of Duisburg-Essen, Campus Essen. Correspondence concerning this article should be addressed to Tobias Kollmann, E-Business and E-Entrepreneurship Research Group, University of Duisburg-Essen, Campus Essen, Universitässtr. 9, 45141 Essen, Germany. E-mail: tobias.kollmann@uni-due.de

REFERENCES

Durand, P. (1983). The public service potential of videotext and teletext. *Telecommunications Policy*, *7*(6), 149–162.

Earston, A. (1980). Viewpoint. *Telecommunications Policy*, *4*(9), 220–225.

eBay (2006). *eBay Inc. Announces fourth quarter and full year 2005 financial results*. Retrieved March 31, 2007, from http://investor.ebay.com/news/Q405/EBAY0118-123321.pdf.

Farrell, J., & Saloner, G. (1985). Standardisation, compatibility and innovation. *Rand Journal of Economics*, *16*(1), 70–83.

Farrell, J., & Saloner, G. (1986). Installed base and compatibility: Innovation, product, preannouncement and predation. *American Economic Review*, *76*(5), 940–955.

Hagel, J., & Armstrong, A. G. (1997). *Net gain*. Boston, MA: Harvard Business School Press.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, *75*(3), 424–440.

Kollmann, T. (1998). Marketing for electronic market places – the relevance of two critical points of success“. *Electronic Markets*, *8*(4), S. 36–39.

Kollmann, T. (1999). Virtual marketplaces: Building management information systems for internet brokerage. *Virtual Reality*, *4*(4), 275–290.

Kollmann, T. (2000). Competitive strategies for electronic marketplaces. *Electronic Markets*, *10*(2), 102–109.

Kollmann, T. (2001). Measuring the acceptance of electronic marketplaces. *Journal of Computer Mediated Communication*, *2*(6).

Kollmann, T. (2005). The matching function for electronic market places – determining the probability of coordinating of supply and demand. *International Journal of Electronic Business*, *5*(3), 461–472.

Kollmann, T. (2006). What is e-entrepreneurship? – Fundamentals of company founding in the net economy. *International Journal of Technology Management*, *33*(4), 322–340.

Lee, H. G., & Clark, T. H. (1996). Impacts of the electronic marketplace on transaction cost and market Structure. *International Journal of Electronic Commerce*, *1*(1), 127–149.

Lumpkin, G., & Dess, G. (2004). E-business strategies and internet business models: how the internet adds value. *Organizational Dynamics*, *33*, 161–173.

O'Reilly, T. (2005). *What Is Web 2.0. Design patterns and business models for the next generation of software*. Retrieved March 31, 2007, from http://www.oreillynet.com/ pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly*, *30*(1), 115–143.

Pavlou, P. A., & Gefen, D. (2005). Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating role. *Information Systems Research*, *16*(4), 372–399.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). New York: Free Press.

Shapiro, C., & Varian, H. R. (1999). *Information rules. A strategic guide to the network economy* (Reprint). Boston: Harvard Business School.

Weiber, R. (1992). *Diffusion von Telekommunikation*. Wiesbaden: Gabler.

KEY TERMS

Critical Mass: Of users is an important success measure for Web 2.0 platforms, because the subjectively perceived attractiveness of a system (e.g. community) is highly correlated with the already registered number of users. A certain number of users within a network are necessary to create value among the participants at a sophisticated level. Reaching this level is essential for a network, because the enrolled participants will be reinforced to use the system on an ongoing basis, and it will become easier to convince new users to join in. The minimum number of participants to maintain a sufficient utility on a long-term basis is referred to as the critical mass.

E-Community: Stands for a virtual community. E-Communities are one sort of communication platform on the internet, and support or initiate business processes. They are used to build constant, self dynamic communication and interaction processes.

E-Marketplaces: Are virtual-based organizations to exchange services. An E-Marketplace has an institutional framework for transaction processes. They can be interpreted as virtual space where supply and demand is coordinated.

Net Economy: Refers to the commercial/business dimension of electronic data networks and is therefore a digital network economy. Different electronic platforms perform the execution of information, communication and transaction processes. The starting point for its expansion is the development of the information society. The basis of the Net Economy is formed by four technological innovations: telecommunication, information technology, media technology and entertainment (the so-called TIME markets). These innovations have, and continue to, significantly impact the possible ways in which information, communication and transactions are managed. The increased support of business processes using electronic systems takes centre stage here. There are a number of terms for this that can be identified (e.g. e-business, e-commerce, information economics, network economics), which can, to some degree, be used synonymously.

Netiquette: Is derived from the terms internet and etiquette. Despite the fact that every human being possesses the right to express an opinion on the internet, legal and ethical standards have to be obeyed. Sometimes those standards are violated, especially on anonymous web-based communication platforms. Comments with an extremist, offending, or sexually harassing content should not be tolerated by any individual or platform operator. Guidance for communication on the internet is given by the so-called netiquette. The recommended behavior of the netiquette is not legally binding, but helps to maintain and develop a positive net culture. The voluntary agreed upon rules are frequently incorporated in codes of conduct of platform operators.

Vapor Marketing: Is characterized by promising announcements to the market on products or services in advance and the management of expectations combined with additional online and offline marketing activities.

Web 2.0: Is the next evolutionary step of the internet. In the past the internet used to be recognized as a technology to publish and distribute data, information and media content. This view was based on split-up roles: Private and commercial publishers of web contents with an active role on the one hand, and passive consumers on the other hand. This golden rule changed in 2005, when Web 2.0 concepts as a new category of websites were established. The traditional differentiation between active content providers and passive consumers diluted. On Web 2.0 platforms users are able to generate and affect contents. User generated content became the slogan of Web 2.0. The active role of users built the basis for innovative business ideas, which were unthinkable before. Many Web 2.0 business models like online communities or video platforms are centered around community structures.

ENDNOTE

[1] Based on Shapiro & Varian (1999, p. 177)

Chapter XXXVII
The Diffusion of WiMax Technology: Hurdles and Opportunities

Phillip Olla
Madonna University, USA

ABSTRACT

A fundamental goal of the wireless industry is to economically offer ubiquitous broadband access to a large number of people in diverse geographical settings. WiMax technology has the potential to be a leading cost-effective technology in the wireless industry, providing opportunities to mobile service providers who lack a Third Generation (3G) license or related infrastructure. WiMax benefits from widespread industry backing and established standards. This chapter will focus on the potential of WiMax technology to deliver personal broadband and fixed Internet capabilities. The chapter applies the diffusion of the Global Internet framework to analyze the current state of WiMax deployments along with the difficulties in developing a WiMax network and building a subscribers base.

INTRODUCTION

Over the last decade mobile voice has evolved from a niche technology to a must have service, with users adopting this innovation into all aspects of their daily lives. In conjunction with the widespread adoption of mobile phones, a broadband revolution is also happening around the globe due to the proliferation of technologies such as Digital Subscriber Line (DSL), cable modem, and broadband wireless services. The next major evolution in communications will see the convergence of both mobile and broadband technologies to create a phenomenon called Personal Broadband (PB) or Mobile Broadband Internet (MBI). Personal Broadband can be viewed as a fusion of the two perpetual markets of mobile voice and broadband, aiming to serve four types of customers: those migrating from mobile voice services and seeking higher speeds for multimedia applications, fixed users who want mobility, WiFi users seeking additional range, and new users who will adopt the new generation of services and applications generated by personal broadband technologies. There are currently three competing technologies capable of broadcasting wireless data at broadband speeds to achieve Personal Broadband. The first is the evolution of the traditional cellular network to provide increasing data bandwidth from a few bits per second with wireless application protocol (WAP) or enhanced data rates for GSM evolution (EDGE) enabled handsets, to a few hundred kilobits per second with evolution-data optimized (EVDO) or high-speed downlink packet access (HSDPA), to the promise of millions of bits per second with the eventual Long-Term Evolution of the 3G standard (3G LTE). Another approach being pursued involves using WiFi to create a mesh network. A wireless mesh network is a network created through the connection of wireless access points installed at each network user's locale. Another interesting approach involves the use of Worldwide Interoperability for Microwave Access (WiMax) technology, WiMax which began as a fixed broadband access technology, and has recently added features to enable mobility. WiMax can deliver capacity similar to the 3G LTE but is being deployed in networks now, while the 3G LTE is not expected to be available for at least two years. This chapter will focus on the potential of WiMax technology to deliver personal broadband and fixed Internet capabilities using the diffusion of the Global Internet framework to analyze the current state of WiMax deployments along with the difficulties in developing a WiMax network and building a subscribers base. The framework employed is beneficial to the discussion because it allows for an analysis along multiple dimensions which incorporate various perspectives such as political, technological, social, economic, and historical factors that have shaped the evolution WiMax phenomenon. WiMax is a standards-based technology that facilitates the delivery of wireless broadband access as an alternative to wired broadband like cable and DSL. WiMax stands for 'World Interoperability for Microwave Access, and is a wireless network infrastructure based on the IEEE 802.16 standard. WiMax provides fixed, nomadic, portable and mobile wireless broadband connectivity without the need for direct line-of-sight with a base station.

This chapter is structured as follows: The first section provides a background discussion on the diffusion of the Internet models along with a discussion on the global diffusion of the Internet framework employed to examine WiMax diffusion; this is followed by an overview of WiMax connectivity infrastructure. The next section describes the market segmentation. This is followed by a discussion on potential hurdles for deploying a WiMax network. Prior to the conclusion, a section discussing the future trend of personal broadband is presented.

BACKGROUND: GLOBAL DIFFUSION OF THE INTERNET THEORY

There are a multiplicity of technology diffusion frameworks, Some frameworks have been adapted to be more specific to the area being discussed such as Internet diffusion in developing nations (Wolcott, Peter et al., 2001), Innovation diffusion in large organizations or specific sectors (Wainwright, David W & Waring, Teresa S, 2007) or to explain mobile data usage (Gruber, H. & Verboven, F., 2001) Mobile data adoption (Fife, Elizabeth & Pereira, Francis, 2005), telecommunication diffusion (Antonelli, Cristiano, 1989). It is beyond the scope of this chapter to provide a detailed review of innovation diffusion research, examples of studies that provide comprehensive reviews are Anand, J. et al. (2006) or McMaster, T & Wastell, D. (2005). There are also interesting studies (Baskerville, R. & Pries-Heje, J., 2001; Kautz, K. & Larsen, E.A., 2000; Mustonen-Ollila, E. & Lyytinen, K, 2003) that make an attempt to create conceptually rich innovation diffusion theory by performing an evaluation traditional organizational setting which leads to greater insight into complex IS adoption and diffusion problems (Wainwright, David W & Waring, Teresa S, 2007).

This chapter is interested in models that consider mobile data services and Internet diffusion, as WiMax technology is a precursor to delivering both mobile data services and Internet connectivity, it is difficult identifying a model to analyze diffusion of a innovative technology on a global scale as most models concentrate on innovation diffusion at the national innovation or organizational level. Researchers must consider choosing from a multitude of frameworks and models, and discover that they must select constructs across the models, or choose a factored model and largely ignore the contributions from alternative models (Venkatesh, V.Ramesh, et.al., 2003). Some researchers believe that the evolution of new models to analyze mobile technology adoption reflects the level of uncertainty faced by industry decision-makers trying to develop investment strategies (Fife, Elizabeth & Pereira, Francis, 2005), however another rationale may be the complexity of the mobile data field and the fact that the field is interwoven with the Mobile Internet adoption.

Some of the frameworks that investigate mobile data services include the model for studying market segmentation of mobile data services (Gilbert, L. A. & Kendall, J., 2003) Another important model is the Input-Process-Output model (IPO) created by Sarker and Wells (2003) which discusses the importance of consumer adoption and usage of mobile devices, Gruber (2003) takes into consideration user characteristics, specific mobile technologies features. Nysveen & Pedersen (2005) created a framework for mobile service adoption that identifies the relationship between external influences, and perceived usefulness. This study also found that attitudes towards usage are influenced by social factors such as media communications along with interpersonal communications (Nysveen, H. et al., 2005). There are existing research into mobile data usage that examine the role of mobile services in a variety of public and private social environments such as, the work and leisure perspective, these are important because they focus on the fact that data usage is undertaken in a variety of different scenarios (Venkatesh, V.Ramesh, et.al., 2003). Other studies use techniques such as demographic surveys, social theory, interview and social observation-based qualitative studies to determine a profile of the mobile user and their use patterns of mobile services (Nysveen, H. et al., 2005). Another interesting theme that has been explored is the link to culture. Culture relates to the way of life, and particularly as the learned behaviour, principles, and beliefs shared by a group of people or a nation (Rosman, A. & Rubel, P.G., 1995). Typically a group of people or society exhibit a range of individual behaviours, however, there are features of personality that a number of people will share along with traits such as norms, language, and shared historical experience. Research has identified national level cultural characteristics (Hofstede, G., 1993), which consist of individualism vs. collectivism, femininity vs. masculinity, long term vs. short term orientation, power distance, and uncertainty avoidance. Prior research has encountered problems in determining the implications of the first four of these dimensions on user adoption of interactive network services. The only dimension that seemed relevant to investigate the diffusion of mobile technology was uncertainty avoidance. Uncertainty avoidance in context of a national trait describes the degree to which a society is open to and accepting of innovations. Consequently, in societies where people have a low uncertainty avoidance there will typically be low resistance to new ideas or technologies (Wainwright, David W & Waring, Teresa S, 2007).

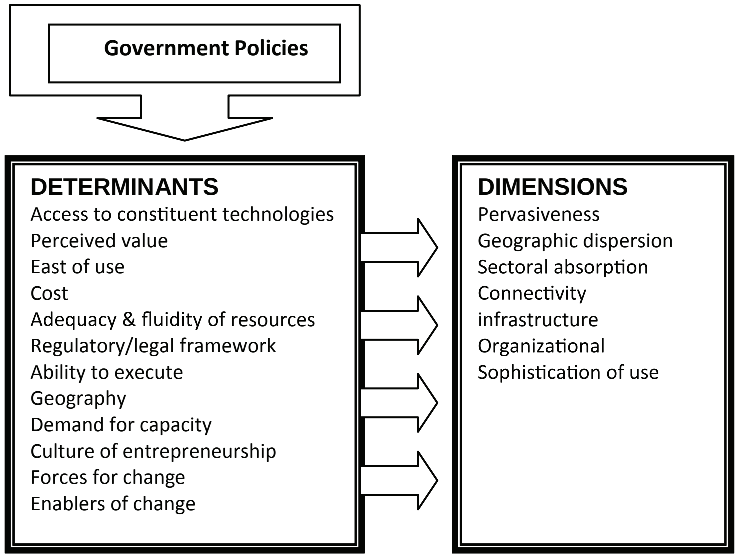
In addition to the mobile data diffusion theme, the other theme that is important to this chapter is the diffusion of Internet studies. The framework used in this chapter is based on an analytic framework for assessing the Global Diffusion of the Internet (GDI) (Wolcott, Peter & Goodman, Seymour, 2003). This framework originated from the Information technology capability of nations study (Wolcott, P., S. et al., 1997), and was later developed into Global Diffusion of the Internet framework (Goodman S, et al., 1998), and has been used to analyze the diffusion of Internet in a variety of countries including China, India (Wolcott, Peter & Goodman, Seymour, 2003), Togo (Bernstein, Adam & Goodman, Seymour, 2005), and Ghana (Foster, William et al., 2004). This framework is a practical resource because it is feasible to measure all values of the variables, provided resources are available. If the analytic framework is based on variables that cannot be measured in practice, then its effectiveness is compromised. The generic framework consists of two important components; dimensions and determinants. Dimensions contain six variables, illustrated in Figure 1; they capture the state of the Internet within a country at a given point in time. While the determinants components reflect the factors that led to the observed state and will likely influence future development. It widely acknowledged that any useful analytic framework should be sufficiently rich that it captures the multifaceted diversity of countries' experiences with the Internet (Foster, William et al., 2004), alternatively, the number of variables should be small enough that they can be easily referenced and utilized. The benefit of this approach for analyzing the global diffusion of WiMax technology for mobile data and internet connectivity is that each variable in the framework describes an important, intuitive, and measurable feature of the presence. Consequently the variables form a comprehensive set in that they collectively comprise of almost everything that might reasonably be of interest, and each variable has something to offer to the complete picture.

DIFFUSION FRAMEWORK FOR WIMAX

This framework presented in Figure 1 below has been adapted in the next section (Figure 2), to account for the fact that, this study is investigating the global diffusion of the Internet and mobile data adoption based on a particular technology instead of assessing generic internet diffusion in a single country. The benefits of this approach become evident when we look at the dimensions and determinant in the next section as it allows for the identification of determinants that are likely to affect future WiMax developments. A brief description of the six dimensions are presented below.

1. **Pervasiveness:** Number of users per capita
2. **Geographic Dispersion:** Physical dispersion of infrastructure and access; primarily a function of the fraction of first-tier political subdivisions (states, provinces, governorates, etc.) with Internet points of presence (POPs).
3. **Sectoral Absorption:** Extent of connectivity in four social sectors: Education, Commercial, Health, and Government.
4. **Connectivity Infrastructure:** Capacity of the technical infrastructure; primarily a function of the capacity of domestic and international backbones, and the types of access (e.g. modem vs. high-speed) available to users.
5. **Organizational Infrastructure:** Internet services market characteristics; a measure of the richness, robustness, and level of choice of the Internet service provision market.
6. **Sophistication of Use:** Integration, transformation, and innovation; a measure of the nature of Internet usage by a leading segment of the user community.

Figure 1: Diffusion of Internet framework (Wolcott, Peter & Goodman, Seymour, 2003)



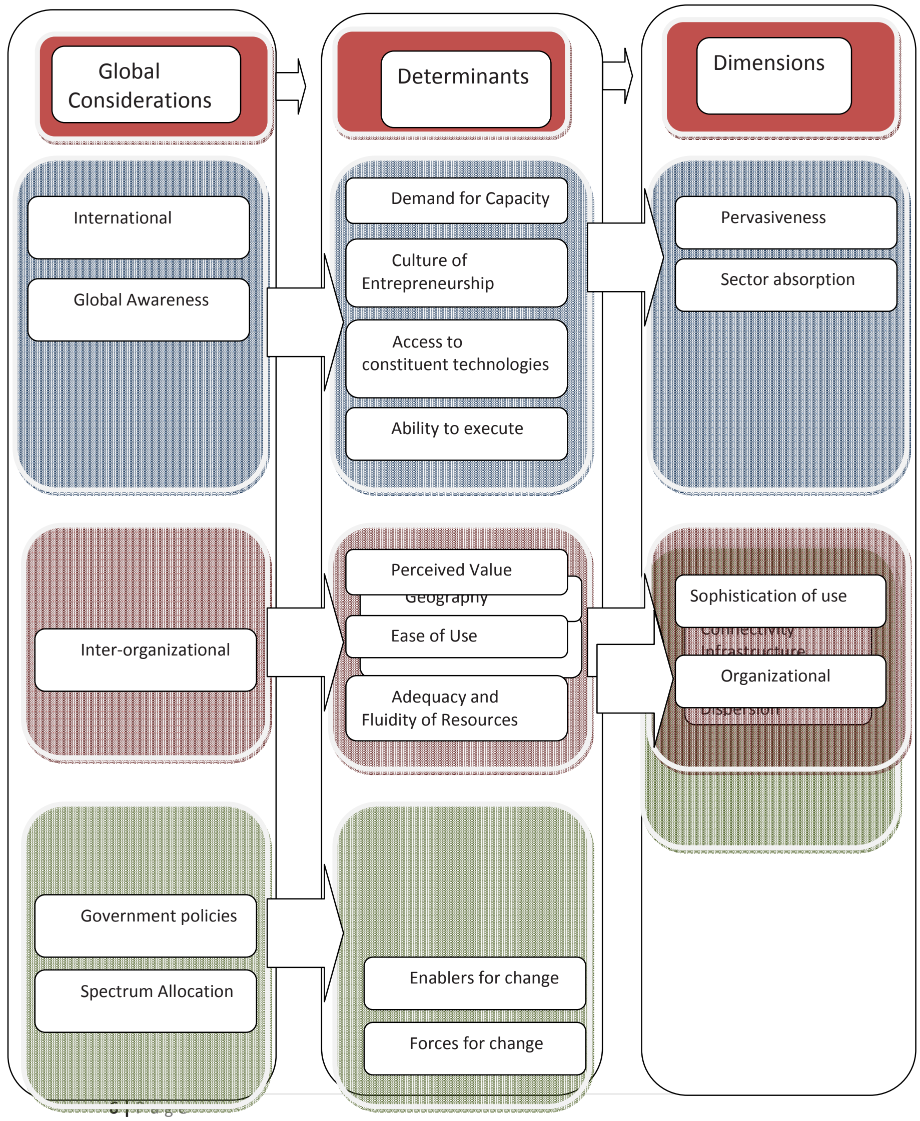
PERVASIVENESS AND GEOGRAPHIC DISPERSION

Compared to competing technologies such as GSM and CDMA, WiMax is considered to be an emerging technology standard and has seen steady growth figures around the world since the standards became ratified in 2004 (Pujolle, Guy, Chaouchi, Hakima, & Gati, Dominique, 2006). There are currently 264 operators around the world being monitored by WiMaxCounts[1]. According to the current figures from WiMax Counts and Maravedis research, approximately 50% of the global WiMax operators are only delivering high speed Internet. The remaining percentage corresponds to operators that are offering different applications, such as Voice Over Internet Protocol (VoIP), Video, Virtual Private Network (VPN). Contrary to belief that most WiMax deployments are trials and tests 65% of operators are already commercial, 14% are trialling, 9% are planning their launch as at quarter two 2008. The breakdown between subscriber types among WiMaxCounts operators was 65% residential and 35% business.

Although WiMax is considered to be the technology that can be used to open up the developing world to an array of Internet services and digital applications, the operator with the largest number of subscribers is currently based in the USA. Clearwire is the largest operator in number of subscribers, with an estimated 443,000 subscribers in the United States at the end of Q1 2008, an increase of 12.5% compared to the 394,000 subscribers reported in Q4 2007. If the USA operators can deploy Mobile WiMax networks there could be a large increase in new subscribers who desire Personal broadband technologies (Fellah, Adlane, 2007). This scenario is dependent on key factors such as timely certification of mobile WiMax equipment, a decrease in Customer Premises Equipment (CPE) pricing, the availability of devices and the emergence of a value added application and services ecosystem.

The USA mobile WiMax network expansions have been held back by a number of issues such

Figure 2. Global diffusion of Internet framework adapted for WiMax



as funding and the lack of a certification. From a funding perspective, Clearwire has merged with Sprint Nextel's (S) Xohm WiMax unit and received 3.2 billion dollars from the cable industry, Google, and Intel, which provides funds and enough spectrum to build a national mobile WiMax network. Clearwire is attempting to sign up 1.3 million subscribers to its WiMax-powered mobile Internet/phone service by the end of 2009, 8.5 million by the end of 2011, and 30.8 million by the end of 2007 (Sprint, 2008).

CONNECTIVITY INFRASTRUCTURE: OVERVIEW OF WIMAX TECHNOLOGY

WiMax was created by the IEEE 802.16 Working Group and their intent is promoted by WiMax forum. The WiMax forum certifies all WiMax equipment to ensure standardisation leads to lower equipment costs. The premise of using WiMax technology is that it will allow Mobile WiMax operators to provide personal broadband value-added services to subscribers at competitive rates. Personal broadband allows a user to access any data or voice application they are accustomed to using from home or the office, along with new innovative applications designed mobility in mind. Network operators will be able to capitalise on the demand for personal broadband among business and consumer users that existing wired and wireless technologies cannot satisfy in a cost-effective way, and they will generate new revenue streams from new services. WiMax has the potential to provide a variety of important services such as:

- Last mile access to residences for basic telephony services
- Providing backhaul for small businesses, enterprises and cell sites
- Provision of services to Wireless Internet Service Provider (ISP or WISP)
- Temporary backhaul for sporting events, concerts and tradeshows in a variety of terrains and geographic locations
- Provision of broadband backhaul for 3G cell sites
- Mobile telephony services using Voice Over Internet Protocol (VoIP)
- Mobile multi-media services based on Internet Protocol (IP)

The technologies and standards behind WiMax were developed by the IEEE 802.16 Working Group. The group first started developing wireless technologies in 2000, publishing the first standard in April 2002 for equipment operating in the 10-66 GHz frequency band (ieee, 2007). This initial range of frequencies required line-of-sight connectivity and large towers, making it more suitable for high bandwidth backhaul. The next task for the group was to extend the standard (IEEE 802.16a) for use in the lower frequency range of 2-11 GHz. The main benefit of this new frequency range was for non-line of sight (NLOS) connectivity. The groups' current standard IEEE 802.16-20043 (previously IEEE 802.16d) deals specifically with wireless connectivity between fixed devices. The most recent and challenging work is the development of a new mobile standard (IEEE 802.16e) that would allow access via portable devices such as laptops, personal digital assistants (PDA) and mobile phones. The fixed and mobile standards have evolved separately due to the complexity of mobile handoffs from one cell to another. Finally, one of the new tasks for the group (IEEE 802.16f) is working on incorporating mesh-networking capabilities into the standard. If it succeeds this could extend the range of networks by allowing each cell in the network to backhaul traffic from other cells, effectively routing around obstacles such as mountains (ieee, 2007).

Fixed WiMax applications are point-to-multipoint thus enabling broadband access to homes and businesses, whereas Mobile WiMax offers the full mobility of cellular networks at broadband speeds. Both fixed and mobile applications of WiMax are engineered to help deliver ubiquitous, high-throughput broadband wireless services at a low cost. Fixed WiMax-certified products based on 802.16-2004 for fixed and nomadic applications are now commercially available and many of the existing fixed WiMax trials will evolve into full commercial deployments in the near future.

Mobile WiMax is based on OFDMA (Orthogonal Frequency Division Multiple Access) technology, which has inherent advantages in throughput, latency, spectral efficiency, and advanced antennae support; ultimately enabling it to provide higher performance than today's wide area wireless technologies. Orthogonal Frequency Division Multiplexing (OFDM) is a technique for transmitting large amounts of digital data over radio waves. OFDM works by splitting the radio signal into multiple smaller sub-signals that are then transmitted simultaneously at different frequencies to the receiver, OFDMA is a multi-user version of the popular OFDM (Johnston, N. & Aghvami, H., 2007). Many next generation 4G wireless technologies may evolve towards

OFMDA and all IP-based networks as an ideal for delivering cost-effective wireless data services.

The main competing technology with OFDM is called Code-Division Multiple Access, (CDMA) is used in both 2G and 3G systems. It is a "spread spectrum" technology, allowing many users to occupy the same time and frequency allocations in a given space. It assigns unique codes to each communication to differentiate it from others in the same spectrum. CDMA enables many more people to share the airwaves at the same time than alternative technologies such as OFDM. However, systems based on CDMA can not consistently deliver the same throughput as those based on OFDM. CDMA is very effective for delivering many simultaneous low data signals, such as mobile voice, but is not well suited for the delivery of high speed signals such as broadband data. OFDM is more effective in handling data. For a more detailed discussion of on performance comparisons between 3G and WiMax (Peng, 2007).

Fixed WiMax has the potential to fill the gap between Wireless LANs and wide area networks. It is anticipated that WiMax-compliant systems will provide a cost-effective fixed wireless alternative to conventional wire-line DSL and cable in areas where those technologies are readily available. In addition WiMax technology can provide a cost-effective broadband access solution in areas beyond the reach of DSL and cable. It is the ideal solution for bridging the digital divide (Serif, Tacha et al., 2007).

Mobile WiMax will fill in the coverage gaps between hotspots and provide nomadic mobility, which is also referred to as Personal Broadband. This allows users with laptops or other devices to stay connected when out of the office, Table 1 describes the mobility access types supported by both Fixed and Mobile WiMax.

The Relationship between Fixed WiMax and Mobile WiMax

The first WiMax Forum-certified products were targeted for fixed WiMax services, not mobile, the main interest and hype from industry has been towards the mobility features provided by Mobile WiMax. A variety of networks operators all over the world are considering WiMax to build national wireless broadband networks (Sprint in the US, Singapore, Taiwan, India).

The IEEE 802.16-2004 standard was designed for fixed-access usage models. This standard may be referred to as "fixed wireless" because it uses a mounted antenna at the subscriber's site. The antenna is mounted to a roof or mast, similar to a satellite television dish. IEEE 802.16-2004 also

Table 1. The different types of WiMax access (adapted from Nokia, WiMax, Intel)

Definition	Devices	Location	Speed	Fixed WiMax	Mobile WiMax
Fixed access	**Outdoor and indoor CPEs**	**Single**	**Stationary**	**Yes**	**Yes**
Nomadic access	**Indoor CPEs, PCMCIA cards**	**Multiple**	**Stationary**	**Yes**	**Yes**
Portability	**iLaptop PCMCIA or mini cards**	**Multiple**	**Walking speed**	**No**	**Yes**
Simple mobility	**Laptop PCMCIA or mini cards, PDAs or smartphones**	**Multiple**	**Low vehicular speed**	**No**	**Yes**
Full mobility	**Laptop PCMCIA or mini cards, PDAs or smartphones**	**Multiple**	**High vehicular speed**	**No**	**Yes**

addresses indoor installations, in which case it may not be as robust as outdoor installations, however it will allow for cheaper customer equipment. The 802.16-2004 standard is a wireless solution for fixed broadband internet access that provides an interoperable, carrier-class solution for the last mile. This technology provides a wireless alternative to the cable modem and can also provide backhaul services, digital subscriber lines of any type. The IEEE 802.16e standard is an amendment to the 802.16-2004 base specification and targets the mobile market by adding portability and the ability for mobile clients with IEEE 802.16e adapters to connect directly to the WiMax network to the standard. These two standards do not directly compete against each other per say, they address different markets and both lead to the creation of unique sets of applications. The main competitor for Mobile WiMax will be 3G network operators, while Fixed WiMax will battle with the Cable and DSL companies. However, Mobile WiMax, operators can offer both mobile and fixed services from the base station, with varying coverage depending on the device being used. This means that a Mobile WiMax base station is more advantageous than a Fixed WiMax and it is important for an operator to consider which technology fits the market they are addressing. Fixed WiMax access will be the early mainstream revenue generator for operators and other service providers. However, Mobile WiMax will provide a much wider choice of applications, such as extending existing office WiFi networks, giving users continuous access when out of the office.

WiMax Performance Comparison

A lot of claims have been made about WiMax performance. The most contentious point is centered on the maximum link performance of 70 Mbps and a range of over 30 miles. Although these figures hold true in some situations, what is generally missed are the realities that these are extremes, and both conditions cannot be met at the same time. WiMax *can* deliver in excess of 70 Mbps under the following conditions – that the link must be good enough to support a high level of modulation; to support this it must be nearby and within the line of sight. The link must also support 2048, in a 20 MHz wide band. It can deliver 2 or 3 Mbps over a range of 30, if there is a line of sight, with no blocking obstacles, and highly directional antennas (Johnston, N. & Aghvami, H., 2007).

SECTORAL ABSORPTION: MARKET SEGMENTATION FOR WIMAX TECHNOLOGY

There is potentially a variety of markets that Fixed and Mobile WiMax will be addressing (WiMax-Forum, 2004). Typically, it takes time for consumers to “buy-in” to a new technology, service or a new provider of that service. In most markets, technical consumers expect new technology, service, and/or provider to be well tested and have an established proven track record before they will sign up for the service. Technologies such as GSM mobile phones and WiFi (IEEE 802.11) have provided ‘credibility’, which has provided a general acceptance of wireless access. Therefore, it is reasonable to anticipate that WiMax technology will have a reasonable adoption rate once the networks are deployed, subscription rates offered by the operator will also have a significant impact on how quickly the technology and services will be adopted. The following sectors are some of the sectors being addressed by the current WiMax deployments:

Residential High Speed Internet Access

Rural subscribers face similar Internet access problems across all the continents (quote). Residential customers in rural areas are limited to low speed dial-up services. In many developing countries there are still regions with no economically viable means for internet access. WiMax technology will enable an operator to economically address this market segment and have a winning business case under a variety of demographic conditions.

Small Office Home Office High Speed Internet Access

This market segment is primarily dependent on the availability of DSL or cable. In some areas, the available services may not meet customer expectations for performance or reliability and/or are too expensive.

Small and Medium Business

This market segment is very often underserved in areas other than the highly competitive urban environment. WiMax technology can cost-effectively meet the requirements of small and medium sized businesses in low-density environments, and can also provide a cost-effective alternative in urban areas competing with DSL and leased line services.

WiFi Hot Spot Backhaul

WiFi hot spots are being installed worldwide at a rapid pace. One of the obstacles for continued hot spot growth however, is the availability of high capacity, cost-effective backhaul solutions. WiMax is an ideal service that can also address the coverage gaps between WiFi hot spot coverage areas (Pujolle, Guy, Chaouchi, Hakima, & GaÃ¯ti, Dominique, 2006).

WiMax Backhaul Capabilities

There are a variety of options for delivering cellular backhaul, the majority of backhaul is done by leasing T1 services from incumbent wire-line operators. With the WiMax technology, cellular operators will have the opportunity to lessen their independence on backhaul facilities leased from their competitors. Outside the US, the use of point-to-point microwave is more prevalent for mobile backhaul, but WiMax can still play a role in enabling mobile operators to cost-effectively increase backhaul capacity using WiMax as an overlay network. This overlay approach will enable mobile operators to add the capacity required to support the wide range of new mobile services they plan to offer without the risk of disrupting existing services.

Public Safety Services and Private Educational Networks

Due to the support for nomadic services and the ability to provide ubiquitous coverage in a metropolitan area Mobile WiMax provides a tool for law enforcement, fire protection and other public safety organizations enabling them to maintain critical communications under a variety of adverse conditions. Private networks for industrial complexes, universities and other campus type environments also represent a potential business opportunity for network operators.

ORGANIZATIONAL INFRASTRUCTURE: HURDLES BUILDING A WIMAX NETWORK

Many factors will influence the success of deploying a WiMax wireless network. Key factors include the availability of spectrum, equipment performance, equipment cost, competition form other operator, availability of applications, infrastructure costs and customer penetration and take up. These are explained in more details below.

Regulation

The introduction of a new set of technologies, such as those underlying WiMax, often have regulatory implications requiring the attention of national regulators. To some regulators, WiMax is a complementary wireless technology to existing 3G, WLAN and wired broadband networks in their country. For others, WiMax is believed to be more of a disruptive technology capable of crossing over and destabilizing existing markets. Irrespective of the perspective taken by policy makers, WiMax and competing wireless technologies could be key components of a future converged network. This means that it is imperative that regulators actively observe developments of the technology and keep abreast of regulatory changes in other countries to encourage innovation and stamp out anti-competitive behavior. Policy makers and regulators will need to ensure that they investigate policy that addresses issues such as: port blocking and traffic structuring, technological neutrality, promoting innovation , and licensing spectrum effectively.

Allocation of Spectrum

This section will describe policy governing the use of spectrum for the equipment behind WiMax and restrictions on spectrum use that could limit services in some markets. WiMax systems ideally require large blocks of spectrum assignments. This enables systems to be deployed with large channel bandwidths, and flexible frequency re-use with minimal spectral inefficiencies for guard-bands to facilitate coexistence with adjacent operators. WiMax Forum is constantly collaborating with standards and regulatory bodies worldwide to promote the allocation of spectrum in the lower frequency bands (< 6 GHz) that is both application and technology neutral. Additionally, there is a major push for greater harmonization in spectrum allocations in order to minimize the number equipment variants required to cover worldwide markets.

The initial system performance profiles that will be developed by the WiMax Forum for the recently approved 802.16e-2005 air interface standard are expected to be in the licensed 2.3 GHz, 2.5 GHz, 3.3 GHz and 3.5 GHz frequency bands. The 2.3 GHz band has been allocated in South Korea for WiBro services based on the Mobile WiMax technology, with a 27 MHz block of spectrum assignment to each operator. The 2.5 to 2.7 GHz band is already available for mobile and fixed wireless services in the United States. This band is also currently underutilized and potentially available in many countries throughout South America and Europe as well as some countries in the Asia- Pacific region. The 3.3 GHz and 3.5 GHz bands are already allocated for fixed wireless services in many countries worldwide and are also well-suited to WiMax solutions for both fixed and mobile services(WiMAX-Forum, 2006).

There is a view that a globally unified spectrum band at 2.4 GHz was one of the reasons for the success of WiFi. The harmonized spectrum band has allowed equipment manufacturers and consumers to benefit from economies of scale, effectively increasing supply and lowering prices for equipment. This has largely been made possible since the equipment has been largely designed for indoor license-exempt use at low power levels. This is not the case for WiMax deployments as spectrum allocations are less harmonized. Technologies under consideration for the first round of testing by the WiMax forum use multiple frequencies to accommodate a wide variety of regulatory regimes.

In many countries the multiple frequency is causing some uncertainty on how the technology could be used in a given spectrum band. For example, the 5 GHz band is open for license-exempt use in some countries but the high frequency ranges make it more suited to fixed broadband access than mobile services. Therefore the 5 GHz ISM band (5.725-5.850) will be ideal for grass-roots wireless ISPs which offer point to multipoint fixed access. The success of WiMax could eventually depend on the ability of operators to find the appropriate and available spectrum. However, without a globally recognized frequency band, the economies of scale may be reduced (WiMax-Forum, 2007a).

On a positive note, the WiMax standards have attempted to alleviate some of these potential problems by restricting the minimum amount of spectrum required for a channel, therefore making it easier to locate bands of spectrum among existing allocations. WiMax channel sizes can range from 1.75 MHz to 20 MHz. In contrast, WiFi requires channel sizes of 20-22 MHz and typically operates in the 2.4 and 5 GHz ranges (WiMax-Forum, 2007b)

There are a number of different markets around the world, depending where you are; the markets differ in terms of frequency, and the need for different services and applications. For frequency, there is something of a coalition around the use of 3.5 GHz frequency for fixed applications of WiMax and 2.5 GHz for mobile applications, but there is also 3.3 GHz available in China and 2.3 GHz in America for fixed applications. In addition, there is new spectrum to be allocated at 1.7 GHz and 2.1 GHz. The 3.3 GHz and 3.5 GHz mobile spectrum is already being used in other parts of the world.

In summary, the market is very fragmented, so according to where the service provider is located and what spectrum is available to them dictates their approach to the adoption of WiMax (McClelland, Stephen, 2006).

Network Equipment Costs

Prior to the advent of WiMax, the high cost of long-distance wireless equipment such as Local Multipoint Distribution Service (LMDS) and ***Multichannel multipoint distribution service*** *(*MMDS) kept the long distant wireless technology as a niche infrastructure. LMDS equipment is approximately USD 4 000 per unit with a base station costing USD 100 000 (Register, 2005). The WiMax Forum aims to keep prices low through standardization in order to push adoption of the technology. It is difficult to estimate costs for equipment before the technology is available but some information is available due to the large number of Pre WiMax trials. Currently, Pre-WiMax equipment is considerably cheaper, with Motorola's subscriber equipment costing less than USD 300. In-Stat estimates that it would cost USD 3 billion in equipment, towers, sites, labor and set-up costs to build a national WiMax network in the United States (Register, 2005) .

Spectrum Costs

The feasibility and financial viability of a network implementation can depend heavily on the costs to use the spectrum. Certain operators are likely to use equipment in license-exempt bands, especially in rural and remote areas. For others who decide to focus on licensed spectrum, costs could be a serious issue that obliterates the project before it even begins. It is important the operators learn from previous spectrum auctions such as the 3G fiasco, which netted governments billions of dollars and left the operators strapped for cash to build the networks along with no discernable business model for recouping the billions of dollars. In Austria for example, WiMax Telecom paid USD 208 000 (EUR 160 000) for a national 3.5 GHz license within a competitive award process (unstrung.com, 2005). In many countries operators are looking to make use of spectrums they acquired for other fixed wireless licenses or earlier mobile radio technologies. Flexibility in national spectrum policies are the key to keeping costs down.

Interference

Some operators are rolling out pre-WiMax equipment in the license-exempt 5 GHz range, specifically 5.8 GHz to avoid expensive spectrum auction scenarios. This may be an inexpensive solution for rural areas with low population densities and little competing uses for the spectrum. However, the use of the license-exempt bands for paid services has some potential downsides. There is a risk that services in license-exempt bands may receive interference from other non-licensed usage, which can slow down or disrupt transmissions. Services offered to businesses and consumers may slow down considerably, or even fail to work, in the presence of new systems deployed by new entrants using the same frequencies. Sometimes being the first provider in the area to use the license-exempt spectrum will create a barrier to entry for the new entrants due to the prospect of interference to their own services from having to share spectrum with the first mover. This issue will be important for regulators who must struggle with conflicting goals of providing initial connectivity and fostering competition (Gunasekaran, Vinoth & Harmantzis, Fotios C., 2007). Allowing the use of unlicensed spectrum for rural network operators serves the purpose of expanding connectivity to areas that were previously underserved.

Interconnection (New Entrants)

Internet Service Providers (ISP) offering WiMax connections will need the ability to interconnect to Internet exchanges, and likely to the PSTN if they are providing VoIP services. Governments have a role to play to ensure that new ISPs are allowed to interconnect to existing networks on similar terms and conditions as other operators. Any fixed-line and mobile operators that built WiMax networks would already have interconnection arrangements in place. However, new entrants to the market would need to be able to pass on Internet traffic and route calls into the PSTN at competitive rates. This means that new entrants would need to negotiate new contracts with the aid of the regulators in their country.

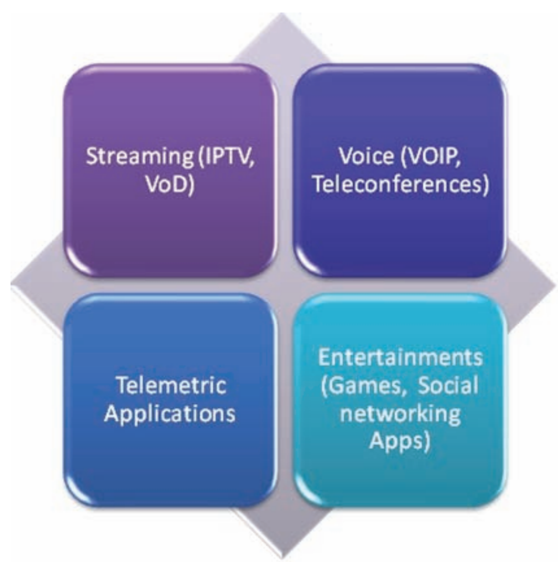
Port Blocking (Not Anti-Competitive in All Countries)

Port blocking occurs when one company prevents certain types of internet traffic from traveling through its networks, the most common cases involves preventing voice subscribers from switching to VoIP or video services Several countries, such as Panama and Egypt have experimented with port blocking as a way to protect the legal national fixed-line monopoly from customer migration to VoIP technologies (CNET.com, 2002) . These efforts have largely failed with VoIP users finding ways around the blockages and a strong negative reaction from the public. Port blocking has recently emerged as an anticompetitive tool by ISPs to prevent consumer access to competing services. This is an important issue for regulators to address. The FCC resolved the port-blocking complaints in the United States but there are more complicated questions that will have to be addressed, as new types services are launched.

SOPHISTICATION OF USE: PERSONAL BROADBAND PARADIGM

The next major evolution in communications will see the convergence of both mobile and broadband

Figure 3. Personal Broadband Applications and Services



technologies to create a Personal Broadband as illustrated in Figure 3.

One of the reasons Personal Broadband will become so important is the fact that consumers are dictating that they stay connected ubiquitously irrespective of their location. Users are now accustomed to broadband at home and expect the same connection to be available in their offices, airports, hotels, and other public spaces, similar to the constant convenience of a mobile phone (Falch, M. & Tadayoni, Reza, 2007). One factor that is driving this trend is the increase in multimedia content available on the internet and mobile networks(Alam, Mahbubul & Prasad, Neeli, 2008). Another important factor is the trends towards Web2.0 applications that require users to connect to the Internet to gain access to services and applications. Consumers are demanding services that are personal, convenient and affordable. The ethos of the 3Ws – 'whatever content', 'Whenever I want It, 'Wherever I am' is a significant shift from the current approach of internet connectivity. The mobile devices that are expected to start this revolution are already available; some examples include the PDA, Smart phones, and laptops (Siegemund, Frank et al., 2005). However, the final key to the puzzle is the availability of regional and national wireless broadband networks such as the US sprint network that will aim to deliver mobile broadband connectivity at fixed costs.

CONCLUSION

This chapter focused on the potential of WiMax technology to deliver personal broadband and fixed Internet capabilities, using components from the global diffusion of the Internet framework to analyze the current state of WiMax deployment utilizing the following six dimensions : pervasiveness, geographic dispersion sectoral absorption, connectivity infrastructure, organizational Infrastructure, and sophistication of use.

Mobile WiMax has the potential to provide mobile users with ubiquitous access, real-time accessibility, seamless interactivity, quality of service, and affordability. These parameters are seen as the new requirements for continued growth and success of the mobile data market and are fundamental for the growth of new applications and services fuelling the Personal Broadband trend.

WiMax utilizes the concept of digitalization, packet transmission, switching, and uses the Internet protocol to support all types of multimedia services such as Voice over IP (VoIP), high-speed internet and video transmission. WiMax will allow network operators service providers to deliver the latest generation of internet services to a user while he/she is at homes, work, Cybercafés, schools or even on the move.

KEY TERMS

3G: The generation of mobile phone standards and technology. It is based on the International Telecommunication Union (ITU) standards under the International Mobile Telecommunications program

CDMA: Code-Division Multiple Access, (CDMA) is used in both 2G and 3G systems. It is a "spread spectrum" technology, allowing many users to occupy the same time and frequency allocations in a given space. It assigns unique codes to each communication to differentiate it from others in the same spectrum.

Fixed WiMax: Fixed WiMax refers to systems built using the IEEE 802.16-2004 ('802.16d') standard and they use the OFDM as the air interface technology. Fixed WiMax deployments cannot cater for handoff between Base Stations, therefore the service provider cannot offer mobility.

Mobile WiMax: Mobile WiMax refers **t**o systems built using IEEE 802.16e-2005 standard and the OFDM as the air interface technology. "Mobile WiMax" implementations can be used to deliver both fixed and mobile services and support hand-off similar to cellular services.

OFDM: Orthogonal Frequency Division Multiplexing (OFDM / OFDMA) technology. This is a technique for transmitting large amounts of digital data over a radio waves. OFDM works by splitting the radio signal into multiple smaller sub-signals that are then transmitted simultaneously at different frequencies to the receiver.

UMTS: Universal Mobile Telecommunications System is one of the third-generation (3G) mobile technologies. Currently, the most common form of UMTS uses W-CDMA as the underlying air interface. It is standardized by the 3GPP, and is the European answer to the ITU IMT-2000 requirements for 3G cellular radio systems.

WiMax: Worldwide **I**nteroperability for **M**icrowave **Ac**cess, is a mobile technology aimed at delivering wireless data over long distances It is based on the IEEE 802.16 standard. The name *WiMax* was created by the WiMax Forum, which was formed in June 2001 to promote interoperability of the standard and certify products.

REFERENCE

Alam, M., & Prasad, N. (2008). Convergence Transforms Digital Home: Techno-Economic Impact. *Wireless Personal Communications, 44*(1), 75-93.

Anand, J., Rottman, J. W., & Lacity, M. C. (2006). A Review of the Predictors,Linkages and Biases in IT Innovation Adoption Research. *Journal of Information Technology, 21*(1-23).

Antonelli, C. (1989). The Diffusion of Information Technology and the Demand for Telecommunication Services. *Telecommunications Policy, 13*(3), 255-264.

Baskerville, R., & Pries-Heje, J. (2001). A Multiple-Theory Analysis of a Diffusion of Information Technology Case. *Information Systems Journal,* (11), 3.

Bernstein, A., & Goodman, S. (2005). Global Diffusion Of The Internet Vi: The Internet In Togo. *Communications of the Association for Information Systems, 15*(1), 371-392.

CNET.com. (2002). *Old-fashioned corporate crackdowns in a brave new world*. Retrieved, from the World Wide Web: http://reviews.cnet.com/4520-3513_7-1027308-1.html.

Falch, M., & Tadayoni, R. (2007). Next generation broadband - Content and user perspectives. *Telematics and Informatics, 24*(4), 243-245.

Fellah, A. (2007). *WiMAX Market Trends and Deployments*. Paper presented at the WiMAX World Europe Vienna, Austria , May 29, 2007.

Fife, E., & Pereira, F. (2005). *Global Acceptance of Technology (GAT) and Demand for Mobile Data Services By*. Retrieved July 4th 2008, from the World Wide Web: http://www.marshall.usc.edu/assets/004/5381.pdf

Foster, W., Goodman, S., Osiakwan, E., & Bernstein, A. (2004). GLOBAL DIFFUSION OF THE INTERNET IV: THE INTERNET IN GHANA. *Communications of the Association for Information Systems, 13*(1), 654-681.

Gilbert, L. A., & Kendall, J. (2003). *A Marketing Model for Mobile Data Services*. Paper presented at the Proceedings of HIICS-36, Honolulu, USA, 89-98.

Goodman S, e. a. (1998). *The Global Diffusion of the Internet Project: An Initial Inductive Study,* . Retrieved July 2008, from the World Wide Web: http://mosaic.unomaha.edu/GDI1998/GDI1998.html

Gruber, H., & Verboven, F. (2001). The Diffusion of Mobile Telecommunications Services in the

European Union. *European Economic Review, 45*(577-588).

Gunasekaran, V., & Harmantzis, F. C. (2007). Emerging wireless technologies for developing countries. *Technology in Society, 29*(1), 23-42.

Hofstede, G. (1993). Cultural Constraints in Management Theories. *Academy of Management Review,, 7*(1), 81-94.

ieee. (2007). IEEE802 Background. *http://www.ieee802.org/16/pub/backgrounder.html*.

Johnston, N., & Aghvami, H. (2007). Comparing WiMAX and HSPA : a guide to the technology. *BT Technology Journal, 25*(2), 191-199.

Kautz, K., & Larsen, E. A. (2000). Diffusion Theory and Practice: Disseminating quality management and software process improvement innovations. *Information Technology and People, 13*(1), 11-26.

McClelland, S. (2006, Tues, May 23. 2006). *WiMAX World Europe: We Want Spectrum, and Good Prices, and, er, LicensesWhat Impediments Are There to WiMAX?* Retrieved, from the World Wide Web: http://www.telecommagazine.com/TechZones/backoffice/

McMaster, T., & Wastell, D. (2005). Diffusion-or Delusion? Challenging an IS Research Tradition. . *Information Technology and People, 18*(4), 383–404.

Mustonen-Ollila, E., & Lyytinen, K. (2003). Why Organizations Adopt Information System Process Innovations: A longitudinal study using diffusion of innovation theory,. *Information Systems Journal, 13*(3), 275-292.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005). Intentions to Use Mobile Services: Antecedents and Cross-Service Comparisons. *Journal of the Academy of Marketing Science, 33*(3), 1-17.

Peng, M., & Wang, W. (2007). A Unified Architecture and Key Techniques for Interworking between WiMAX and Beyond 3G/4G Systems. *Wireless Personal Communications.*

Pujolle, G., Chaouchi, H., & GaÃ¯ti, D. (2006). A global architecture for the Wi-family. *Telecommunication Systems, 31*(2), 163-173.

Pujolle, G., Chaouchi, H., & Gati, D. (2006). A global architecture for the Wi-family. *Telecommunication Systems, 31*(2), 163-173.

Register. (2005, 14 February 2005). *WiMAX turns the screw on 3G., The Register,* . Retrieved, from the World Wide Web: http://www.theregister.co.uk/2005/02/14/wimax_versus_3g/print.html.

Rosman, A., & Rubel, P. G. (1995). *The Tapestry of Culture: an introduction anthropology, .* , New York: McGraw-Hill Inc.

Sarker, S., & Wells, J. (2003). Understanding Mobile Wireless Device Use and Adoption,. *Communications of the ACM, 46 (12). 46*(12).

Serif, T., Ghinea, G., Stergioulas, L., Chen, S., Tiropanis, T., & Tsekeridou, S. (2007). Satellite-based delivery of educational content to geographically isolated communities: a service based approach. *Personal and Ubiquitous Computing.*

Siegemund, F., Floerkemeier, C., & Vogt, H. (2005). The value of handhelds in smart environments. *Personal and Ubiquitous Computing, 9*(2), 69-80.

Sprint. (2008). *Sprint and Clearwire to Combine WiMAX Businesses, Creating a New Mobile Broadband Company*. Retrieved, from the World Wide Web: http://newsreleases.sprint.com/phoenix.zhtml?c=127149&p=irol-newsArticle_newsroom&ID=1141088

unstrung.com. (2005). *WiMAX Telecom Eyes Standards*. Retrieved, from the World Wide Web: http://www.unstrung.com/document.asp?doc_id=66413

Venkatesh, V. R., et.al. (2003). Understanding Usability in Mobile Commerce. *Communications of the ACM, 46*(12), 53-56.

Wainwright, D. W., & Waring, T. S. (2007). The application and adaptation of a diffusion of innovation framework for information systems research in NHS general medical practice. *Journal of Information Technology, 22*(1), 44-58.

WiMax-Forum. (2004). *Business Case Models for Fixed Broadband Wireless Access based on WiMAX Technology and the 802.16 Standard*. Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads

WiMAX-Forum. (2006). Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation.

WiMax-Forum. (2007a). *egulatory Position and Goals of the WiMAX Forum*. Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads/WiMAX_Forum_Regulatory_Whitepaper_v06162004.pdf.

WiMax-Forum. (2007b). *IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access. available at.* Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads/WiMAXWhitepaper.pdf.

Wolcott, P., & Goodman, S. (2003). Global Diffusion Of The Internet I: India: Is The Elephant Learning To Dance? *Communications of the Association for Information Systems, 11*(1), 560-646.

Wolcott, P., Press, L., William McHenry, Seymour E Goodman, & Foster, W. (2001). Framework for Assessing the Global Diffusion of the Internet",. *Journal of the AIS,, 2*(5).

Wolcott, P., S. , Goodman, & Burkhart, G. (1997). The Information Technology Capability of Nations: A Framework for Analysis,. *MOSAIC Group Report.*

ENDNOTE

[1] WiMaxCounts is an online interactive Operator Tracking Service database profiling WiMax operators worldwide on a quarterly basis .

Chapter XXXVIII
VoIP Quality and Security Issues for Consumers and Small Businesses

David Bell
Pacific Union College, USA

Sunil Hazari
University of West Georgia, USA

ABSTRACT

VoIP is the use of Internet protocols to provide telephone services that have previously been delivered over traditional telephone networks. Advantages of VoIP include cost, portability, and functionality which are the main reasons that many consumers and small businesses are considering this technology as a replacement to traditional telephone services. There are however risks associated with VoIP services which impact quality and security of the phone system for voice communications. This chapter reviews issues related to quality and security as faced by consumers and small businesses. Recommendations are provided to improve call quality and mitigate threats faced in the VoIP environment.

INTRODUCTION

In recent years, technologies have evolved that make it much easier for consumers and businesses to replace the use of the traditional Public Switched Telephone Network (PSTN) with alternate voice communications systems, including Voice over Internet Protocol (VoIP). Many are switching to VoIP to reduce costs and to allow integration of voice with other web applications (Rash, 2005). There are competing technologies from various companies such as Skype and Vonage that each have advantages and disadvantage in terms of costs, interoperability, functionality, and quality (Mark, 2005).

VoIP usage has increased dramatically in recent years. In Q1 2006, 3.9 million U.S. households used VoIP for their home phone service, accounting for 3.5% of the U.S. consumer market (Frommer, 2006). Nearly 48% of VoIP subscribers used Vonage as their service provider, with its closest rival Skype holding 12% of the VoIP market. By Q4 2007 the number of subscribers more than tripled to over 13 million households (Burton, 2008). However, Comcast and Time Warner, two large cable companies that also bundle VoIP service, now command over half of VoIP subscribers with Vonage falling into third place at 20% and Skype falling into a distant fifth place at 8%. Together, combined quarterly revenue of the top six VoIP providers now exceeds $1.4 billion (Elliott, 2008).

Regardless of the particular VoIP brand or protocol utilized, there are several significant factors that must be considered before replacing PSTN service with a VoIP-based solution. Some of the most important problems with VoIP technology include reduced call quality (Wilson, 2005), reduced privacy of communications, and the inherent lack of security found on the public networks that deliver VoIP data (Garretson, 2005).

Although VoIP technologies provide many opportunities for reducing cost and adding functionality not readily available from PSTN solutions, there are a number of factors affecting quality and security of such communications that must be explored to determine their impact on the viability of using VoIP for day-to-day voice communications. The objectives of this paper are to explore methods for improving VoIP call quality and security issues, primarily in the context of consumers and small businesses. This paper focuses on issues affecting call quality that are related to encoding techniques, transport methodologies, and network traversal. The quality of specific hardware, such as microphones, earpieces, routers, and switches are not considered. The security issues examined are limited to the privacy of the content of the phone conversations and possible network intrusion vulnerabilities introduced by the implementation of VoIP.

BACKGROUND

The concept of utilizing the Internet to transmit voice conversations is not a new one. The idea dates back to the 1970s, though the data rates and connectivity available to most individuals made it impractical for common usage at the time (Schulzrinne, 1999). As Internet access proliferated in the mid 1990s, companies such as Vocaltec emerged selling software to facilitate VoIP conversations (Varshney, Snow, McGivern, & Howard, 2002). The products generally required both parties in a conversation to be running the same proprietary software on their computers, and did not provide a path to the PSTN, severely limiting their usefulness for general voice communications.

Since that time, VoIP technologies have continued to evolve, including establishment of several important standards. Development of the Session Initiation Protocol (SIP) has helped to standardize procedures for establishing, changing, and terminating of Internet-based phone calls (Schulzrinne & Rosenberg, 1998). Development of *codecs* for encoding of the audio signal based on available bandwidth and capabilities of the hardware and software at each end point (Garbin & Gharakhanian, 2006) has facilitated communications between software and hardware of multiple vendors, as well as connections into the PSTN (Varshney et al., 2002). As high speed Internet access has continued to become more ubiquitous, an increasing number of businesses and consumers have realized the significant cost savings that can be realized by switching some or all of their voice services to VoIP.

Several questions are addressed in this study. With the increasing availability and adoption of VoIP technologies by both consumers and businesses, there are concerns in the areas of quality, privacy, and security when compared to the traditional PSTN. This paper will explore factors of VoIP implementations that affect communication quality, and methods that may be utilized to mitigate these factors. Threats to the security of VoIP communication and protections against those threats will be also be explored.

VoIP Call Quality

Several studies have shown that for significant amount of the time, call quality for IP-based calls has been unsatisfactory. As an example, Wilson (2005) noted that in a study over a two-month period, Seattle-based InterNAP, a network optimization technology company, simulated 1.5 million VoIP calls from 15 different locations over the backbone of multiple carriers such as AT&T, Qwest, and Sprint, and found that calls were of acceptable quality only 60% of the time, unless some additional efforts were made to optimize the network connection. Call quality has been traditionally measured using factors of clarity, delay, and echo (Garbin & Gharakhanian, 2006). Clarity refers to the clearness or lack of distortion heard by the person on the receiving end of a conversation. Delay measures the time it takes for sounds entering the microphone at one end of a conversation to appear at the speaker or earpiece at the other end. Echo refers to an image of the original sounds bouncing back to the earpiece at the source. Echoes can be perceptible if the round-trip delay exceeds 30 milliseconds (Laurence, 2005). Traditionally, these quality factors have been combined into an overall call quality score called the Mean Opinion Score (MOS) which is based on subjective scoring of voice samples on a 5-point scale (International Telecommunications Union, 1996). Since subjective scoring by human listeners is both costly and difficult to repeat, which makes it impractical for real-time evaluation (Rubino, Varela, & Bonnin, 2006), there are a number of objective algorithms that have been developed to replace the subjective MOS scale by attempting to compare the waveform of the received audio with the original. The ITU has developed the Perceptual Evaluation of Speech Quality (PESQ) model in an attempt to objectively map measurements of the one-way signal distortion to the MOS scale (Garbin & Gharakhanian, 2006). Rubino et al. (2006) however, contend that the metrics used to measure perceived quality based on purely objective factors do not correlate well with real subjective assessments. They propose a hybrid approach that trains a neural network with results of a combination of subjective and objective measurements of a set of degraded audio samples so that the model can learn the relationship between the parameters that cause the distortion and the perceived quality.

Distortion on VoIP calls can be caused by a number of factors. The analog-to-digital conversion, and digital-to-analog conversions that are required to send the voice data over the digital network utilizes one of several available coder/decoder (codec) schemes. The most common codecs used in VoIP include G.711 and G.729a, which allocate 64kbps and 8kbps of bandwidth respectively (Garbin & Gharakhanian, 2006). Even under the best of circumstances, the G.711 and G.729a codecs can only provide MOS scores of 4.4 and 4.2 respectively (Brix Networks, 2007). Measurements by Garbin and Gharakhanian (2006) contradict these estimates primarily for the G.729a codec which was found to have an average MOS score of 3.6 in a controlled IP-to-IP network environment, with much wider variability using different voice samples than for G.711 results. Performance dropped even further if one end of the conversation terminated at the PSTN. The Skype service uses a proprietary codec that is not publicly available, so the codec's affect on distortion is unknown, but the company claims codec bandwidth varies between 3 and 16kbps, which is significantly lower than the G.711 codec (Skype, 2007). It should be noted that using a proprietary codec keeps Skype's users from directly communicating with users of other protocols without additional decoding/encoding processes, further increasing distortion.

Another factor that can cause distortion is the loss of data packets between source and destination. Network congestion, whether due to limited available bandwidth at source or destination, or due to network traversal problems, can lead to packet loss that will distort the received audio (Karol, Krishnan, & Li, 2005). In a time-sensitive application such as VoIP, even packets that are not lost but only delayed may arrive too late to be integrated into the real-time audio stream. Distortion can also result from additional transitions between analog and digital domains if a portion of the call path traverses the PSTN. Each domain conversion, as well as noise introduced in analog segments of the call route, further degrade the waveform of the received audio as compared to the original (Garbin & Gharakhanian, 2006). Delay in VoIP call audio can be attributed, in part, to the process of encoding the audio, "packetizing" the encoded data, processing, queuing, and variations in buffering (Laurence, 2005). According to Laurence (2005), these factors alone can cause delays of 30 to 110 milliseconds even within a local area network environment. When VoIP calls are routed over the public Internet, data packets are subject to latency as they traverse through multiple routers and paths along the way to their destination, further increasing total delay (Garbin & Gharakhanian, 2006).

Echo can occur at several points in the call path. The earpiece speaker audio at the receiving end of the call can be picked up again by the microphone and retransmitted back to the person speaking (VoIP Mechanic, 2006). Echo can also occur when the circuitry within the receiving device has an impedance mismatch that causes a reflection of the received signal back over the two-wire phone connection (Hunt & Arden, 2005). Impedance mismatches can also occur at any analog connection point along the connection path, possibly causing multiple echoes (Davidson, Bhatia, Kalidinidi, Mukherjee, & Peters, 2006).

The primary methods noted for reducing distortion in VoIP calls include choosing a higher quality codec such as G.711 rather than G.729a or other lower bandwidth codecs (Garbin & Gharakhanian, 2006) , and making improvements in the delivery of the data packets in a timely manner to the destination location (Karol et al., 2005). While few alternate codec options were discovered, there are several methods proposed for improving the packet delivery. One solution proposed by several authors such as Galbraith et al., (2005), Rubino et al, (2006), and Edwards (2007) involves implementation of Quality of Service (QoS) mechanisms to give VoIP data packets priority over other types of traffic, reducing the delivery delay and reducing the probability of lost packets. At the LAN level, hardware is readily available that will prioritize time-sensitive traffic such as VoIP, often to the detriment of delivery of other traffic types (Rubino et al., 2006). At the present time there appears to be few options for prioritizing traffic over the Internet since the current state of the public Internet is that it is a "best effort" service treating all packets equally. As the next version of the IP protocol (IPv6) becomes more widely deployed, it will add additional header information that will allow for QoS packet handling (Huston, 2007). Yoo (2006) contends that the current status of "Net Neutrality" in handling all packets equally needs to be changed to a prioritization model, including allowing the market to determine variable pricing structures not only for bandwidth, but for quality of service as well. One distinctly different method of improving packet delivery and reducing delay is proposed by Karol et al. (2005). Devices are installed at sending and receiving locations that will duplicate each outgoing packet and route the duplicate over a different path. The first of the two identical packets to reach the destination device is delivered, and the packet arriving later is discarded. A sliding-window buffer is used for identification of duplicate packets to be discarded.

Packet delay also has an impact on the echo that is heard by a person speaking over a VoIP connection. There are several methods for dealing with this echo. While echoes less than 30ms are generally imperceptible to the person speaking, VoIP communications will almost always require some method of echo cancellation since one-way latency with VoIP will always exceed this threshold (Laurence, 2005). The first echo reduction technique is to simply reduce the volume of the speaker in the handset or earpiece or reduce the sensitivity of the microphone so that the received sound is not retransmitted back to the person speaking (VoIP Mechanic, 2006). Other methods include installation of echo cancellation algorithms in IP-PBX systems and ATAs (the bridge between a customer's telephone line and their ISP modem) that analyze the voice data, and inject a portion of the inverse of the incoming waveform onto the outgoing signal, thereby cancelling the echo. These systems must be able to analyze the amount and timing of signals that are echoing back from the receiver to determine the amount of cancelling signal to inject (Laurence, 2005). Finally, utilizing equipment that has proper impedance matching within analog portions of the signal path will reduce echo (Hunt & Arden, 2005). One quality improvement initiative on the horizon that is independent of the actions of VoIP users involves a partnership between VoIP service providers and hardware manufacturers to monitor the quality of VoIP communications by incorporating capabilities into the ATAs and IP-PBXs to forward real-time performance information to a "service assurance" company for analysis (Wilson, 2007).

VoIP Vulnerabilities to Privacy and Network Security

Unlike PSTN networks which are relatively difficult to tap without physical access to private phone company lines, VoIP communication may be subject to eavesdropping by anyone on the same LAN, or with access to the data stream through a shared medium such as neighborhood cable modem service (Garretson, 2005). If the VoIP user connects through an open wireless hotspot, there is also an increased risk of eavesdropping, and limitations in the SIP protocol may allow the "hijacking" of SIP registration, with a hacker able to divert calls to a different location (Buckler, 2006). VoIP is also susceptible to Denial of Service (DoS) attacks on the network which choke available bandwidth and may make VoIP communications impossible during the attack (Garretson, 2005). Several authors such as Buckler (2006), Erlanger (2005), and Garretson (2007) note the possible problem of Spam over Internet Telephony ("SPIT") which may allow spammers to send huge quantities of voice messages to the inboxes of VoIP phone users since many are accessible via an Internet address and do not require the use of PSTN. Much like e-mail spam, it costs little to send the same voice message to many VoIP users. Another threat noted by several authors is the difficulty of properly configuring firewalls to allow for VoIP traffic without weakening the security from hackers attempting to break through to the internal network. One CTO cited by Buckler (2006) indicated that due to the nature of the SIP protocol's dynamic assignment of ports, he had to open a 10,000-port range of addresses in the firewall to accommodate the VoIP traffic, opening up the network to intrusion. The Skype software includes port cloaking features that can make it easier to traverse NAT devices and appear to be data bound for a Web server, causing difficulties for network managers attempting to configure firewalls to block such traffic (Morrissey, 2005).

Recommendations for Quality Improvement

As millions of users are lured away from traditional PSTN systems to VoIP systems due to lower costs and in increased feature sets, several problems must be kept in mind. VoIP is not the same as PSTN in several important ways. The quality, at best, can barely match a PSTN long-distance call, but it can frequently drop to cell-phone quality, or in some cases not work at all. Different codecs used by different service providers and hardware vendors provide differing levels of quality due to analog-digital conversions and compression schemes. Packet loss and delay can impact the quality of the conversation as well.

There is consensus in the literature that the first stage of call degradation occurs at the point of analog-to-digital conversion which occurs at either the computer softphone, inside an IP phone, or within the Analogue Terminal Adapter (ATA) (voip-info.org, 2007) that connects a traditional telephone to the IP network. The particular codec chosen, whether it is G.711, G.729a, or another protocol such as the one utilized by Skype, can significantly impact the quality of the audio that reaches the other end. If possible, the G.711 codec should be used due to its greater preservation of the original call quality. However, if there are multiple applications accessing a shared connection to an ISP the packet loss due to lack of bandwidth (particularly for outbound data from customers on an asymmetric connection) may make this codec unusable. In such cases, the G.729a codec may be used to make a call feasible, though of inferior quality. In the absence of empirical data on Skype protocols and performance, users of such proprietary codecs will need to test the quality for themselves to see if it provides an acceptable cost/performance ratio. It should be noted that while PSTN also uses G.711 encoding, Garbin & Gharakhanian (2006) indicate that using this codec over an IP network adds additional bandwidth requirements due to the data that must be added to each packet to route it through the IP network. This additional overhead increases bandwidth requirements to 82kbps for G.711 and 26kbps for G.729a. In many cases, the choice of codec is not up to the individual user, but depends on the specific device, software, and service used to perform the translation. For example, one VoIP provider, ViaTalk, only supports G.711, so a user that wishes to utilize the lower bandwidth of G.729a would not be able to use their service (Broadband Reports, 2007). For readily available codecs, G.711 provides the best encoding/decoding quality provided that sufficient bandwidth is available. If there is a choice of codecs, network bandwidth can affect call quality.

Packet loss and/or delay due to network congestion, limited available user bandwidth, and link or router troubles will have a detrimental effect on call quality, with the potential to reduce quality to unacceptable levels. Several authors promote improvements to packet delivery by adding QoS information to packets to give them priority over less time-sensitive data. While this may be appropriate for VoIP over company-controlled networks utilizing QoS-capable switches and routers, none of the authors indicated that this would be a viable solution for VoIP that traverses the Internet until IPv6 becomes widely implemented. At present, increasing user bandwidth to a higher level of DSL or Cable Modem service may decrease the likelihood that multiple applications will interfere with a voice conversation, but it is no guarantee that the packets will not encounter congestion or outages along the path to their destination that affects the quality of the conversation, or terminates it entirely. As more consumers use their broadband connection for audio, video, and other bandwidth-intensive applications, QoS within the home network will become more important (Rubino et al., 2006).

The proposal by Karol et al. (2005) to duplicate VoIP packets over two independent networks reduces packet delay and loss, however it requires specialized equipment at both ends of the network path, and also incurs additional data transmission costs. Though the method may be effective for voice communications between multiple sites of the same company, or for extranet use, there is no indication that it is viable for consumers and small businesses wishing to use VoIP primarily to communicate with PSTN users or other VoIP users at multiple locations.

There is consensus in the literature that, unlike short-distance PSTN connections, echo must be dealt with in VoIP conversations regardless of the physical distance between the participants due to latencies introduced by encoding, packetizing, routing, and decoding of the voice data (Laurence, 2005). Without addressing echo issues, any caller would have a noticeable delayed echo of their own voice coming back into their earpiece or handset. The most important and simplest of solutions is to ensure that the audio from the earpiece not be so loud as to be picked up by the microphone and retransmitted. Using a quality telephone with proper impedance characteristics that match the device that will be sampling the analog audio will also reduce echo. Impedance mismatches that occur within analog portions of the route between the termination points may cause echo that will be much more difficult to remove. If the problem is ongoing, the service provider may need to be contacted for possible solutions. Many ATA devices contain echo suppression algorithms that have user adjustable settings that can significantly reduce echo, but the improvement will be heard by the individual on the opposite end of the conversation rather than by the one attempting to adjust the ATA. Making effective adjustments to these settings requires someone at the other end of the line to provide opinions on the relative effectiveness of different parameter combinations until echo is reduced to a satisfactory level.

Privacy and Security Recommendations

VoIP conversations are subject to eavesdropping unless certain steps are taken to protect them. If an eavesdropper has access to the data packets, software is readily available to capture and decode the Real-time Transport Protocol (RTP) packets that contain the voice data, reassemble them, and save them as an audio file (Thermos, 2006). Eavesdropping within a company network may be mitigated somewhat through the use of VLAN technologies that limit access to the data packets (Erlanger, 2005). Erlanger (2005) also notes that some vendors recommend that companies separate VoIP traffic from other traffic, though this would require duplicate network cabling and hardware and negate much of the cost reduction that led to the implementation of VoIP in the first place. Rash (2005) suggests that firewalls used should be SIP-aware—able to identify packets at the application level so that open ports for VoIP communications cannot also be used for other network data transfer. For business VoIP solutions, additional security may be enforced through the prohibition of installation of VoIP communications that use a headset attached to a computer. Such "softphones," including Skype's client software, and x-lite, may be blocked from downloading and installation through domain policies and firewall filtering (Garretson, 2005), though many end users are savvy enough to bypass such filtering techniques.

While corporate VoIP users may be able to encrypt traffic between locations that they control, consumer and small business VoIP users at present must have little expectation of privacy of their conversations that travel through the Internet. Few users encrypt e-mails, so for many, even though it is possible for someone to intercept their VoIP call, the low probability of having an eavesdropper on any particular call is sufficient peace of mind for millions of users, at least for now. If communications need to be secret, encryption techniques must be used, however with currently available systems, it appears that the encryption technologies from different vendors may be incompatible and may require the same hardware at both ends of the conversation, severely limiting the number of different locations that may be contacted securely (Rash, 2005). There is a Secure Real-time Transport Protocol available (SRTP), but again, there are discrepancies between implementations from different vendors (Mier, 2007). At present there is no indication that consumers and small businesses that use low-cost

VoIP service providers can have any expectation of privacy for their phone calls.

After the limited privacy protection, the security issue that appears to be most worrisome is the susceptibility to intrusion into the network that VoIP configuration can cause. Firewalls that cannot inspect packet contents for VoIP protocols are difficult to harden to intruders if there is a need for internal users to access external VoIP services. Opening ports to allow for SIP packets to traverse the firewall may allow for easier intrusion by hackers probing security holes in the corporate network. Opening firewall ports and configuring packet sniffers to allow for business VoIP communications may be an inconvenience for the corporate network administrator, but there is little evidence that using VoIP services from a home computer has led to significant increases in intrusions. Some VoIP providers advise placing their ATA as the first device connected to the DSL or Cable Modem, and the rest of the home be network connected on the LAN side of the ATA (ViaTalk, 2007). With such a configuration, the home network has increased protection from attack since the ATA provides additional firewall functionality. Optimally, firewall configuration should allow for inspection of packet contents so that ports can be opened for voice traffic but blocked for all other services (Rash, 2005).

Business users should be especially wary of the Skype software client since it can utilize port 80, pretending to be normal HTTP web traffic, and easily traverse NAT devices making it much more difficult to block (Morrissey, 2005). Employee training, and restrictions on downloads and client installation should help mitigate its proliferation. Of course, some businesses and many home users wish to use Skype, and should carefully monitor user forums and industry discussions for any threats that may be discovered in this proprietary technology.

Finally, if users are completely reliant on VoIP for their voice communications, they leave themselves vulnerable to communications disruption due to network outages, denial of service attacks, and power outages. Fallback communication methods must be available in the case of one of these disruptions.

FUTURE TRENDS

The VoIP landscape continues to evolve quickly, with significant legal, technical, and financial issues affecting call availability, quality, and support. As an example, SunRocket, one of the larger VoIP providers went out of business without warning in July 2007, leaving over 200,000 customers scrambling for alternative phone service (King, 2007). Many customers had prepaid for two years of service, and were left with no recourse but to sign up with another company. Other VoIP providers such as ViaTalk and 8x8 have taken over many of the SunRocket contracts (Scoblete, 2007) but have had difficulty handling the influx of new customers. Increased network traffic, shortages of hardware adapters, and long customer service hold times have impacted existing customers as well as those trying to find a replacement for SunRocket

Patent lawsuits from Verizon and Sprint may eventually cost Vonage, the largest independent VoIP provider, well over $200 million (Leslie, 2007), threatening its very existence. If these lawsuits hold up on appeal, it may spell doom for not only Vonage, but also for smaller VoIP providers that are using similar technologies. At the same time, the large phone and cable providers are offering bundled Internet, TV, and Voice services. Even though the VoIP offerings from such companies as Comcast are currently priced approximately three times higher than smaller independent providers such as ViaTalk ($33 per month versus $12), it appears that many customers have decided that one-stop shopping and company stability are more important than rock-bottom pricing (King, 2007). Mergers, buyouts, and vendor consolidations are likely to continue, giving a smaller number of providers access to increasing numbers of customers (Kretkowski, 2006).

On the technology front, increasing rollouts of broadband connectivity, including significantly higher speeds offered by newer options such as Verizon's 30Mbps FiOS service, rolling out in 16 states (Bertolucci, 2007), along with increasing Internet backbone capacities promise to provide more reliable delivery of VoIP packets, improving call quality. VoIP service providers are also learning to tune their systems to dynamically respond to changing Internet conditions. Having multiple SIP servers spread across the country reduces latency in calls. In addition, companies are working to improve automatic reprovisioning of customer's ATA device to switch to a different server if there is a network bottleneck or server failure. Although the concept of such *dynamic routing* for VoIP has been around for several years (Kong & Mase, 2003), companies are still working to improve implementation of these fail-over procedures to help ensure that customers' voice packets are following the fastest, most reliable path on each call.

Wireless VoIP also shows promise as hardware vendors are making cell phones with Wi-Fi capability. It has recently become possible to initiate a VoIP call through a public wireless hotspot, and seamlessly transition to a cell tower connection as the user walks down the street. Customers can now use the same phone to connect to their home wireless router as they use while in their cars. While technically possible today, cellular providers have a long history of disabling certain features of phones they sell (such as direct transfer of ringtones from computer to phone), and it may take some time before they develop a pricing model that allows them to recoup revenues lost if they were to allow customers to use wireless access points instead of plan minutes. In any case, as whole cities are rolling out wireless Internet access, development of phones capable of communicating with multiple protocols seems a natural evolution.

CONCLUSION

In the age of digital convergence, VoIP should be considered a serious contender. When deploying VoIP, consumers and small businesses should study the impact of call quality as well as security. Both of these parameters can impact the projected quality of a service as perceived by the entity receiving the call at the other end. As described in this paper, distortion of call quality can be caused by a number of factors but these factors can be overcome by some simple analog echo reduction techniques and by providing quality of service mechanisms that improve the timely delivery of VoIP data packets. Interoperability between multiple VoIP providers is lacking due to proprietary codecs being used by vendors, but the vendors are beginning to realize the benefits of convergence and open communication between systems. Security and privacy are other significant concerns for VoIP conversations. It is possible to encrypt VoIP data to ensure privacy of voice calls, but it is currently impractical for use on VoIP to PSTN calls and calls between different VoIP provider systems. Countermeasures such as stateful firewalls can be implemented to mitigate threats that are specific to VoIP traffic. There are other disadvantages associated with this technology: VoIP services may not operate in the event of a power outage. Also, 911 emergency services may not be available when using VoIP phones. Despite these problems, VoIP brings many advantages to the consumer. The advantages of cost reduction, increased flexibility, unified communication, extra feature set, automatic routing, and portability are main drivers for its success. All these factors should be considered carefully when selecting VoIP as a primary (or supplemental) telephone service. With identification of factors that impact VoIP and its associated limitations, rapid changes regarding quality, security, reliability, and regulatory issues are poised to make this technology a viable option for consumers and small businesses.

REFERENCES

Bertolucci, J. (2007). Broadband EXPANDS. *PC World*, *25*(8), 76-88.

Brix Networks. (2007). *Where does the 1-5 quality scale come from?* Retrieved November 24, 2007 from http://www.testmyvoip.com/faq.html#mos

Broadband Reports. (2007). *CODEC Choice in ViaTalk*. Retrieved November 27, 2007 from http://www.dslreports.com/forum/remark,18142109

Buckler, G. (2006). VoIP systems bring in new vulnerabilities. *Computer Dealer News, 22*(9), 26.

Burton, T. (2008). *Q4 consumer VoIP league table*. Retrieved March 9, 2008 from http://www.fiercevoip.com/story/q4-consumer-voip-league-table/2008-03-03

Davidson, J., Bhatia, M., Kalidinidi, S., Mukherjee, S., & Peters, J. (2006). *Voice over IP fundamentals* (2nd ed.): Cisco Press.

Edwards, J. (2007). *VoIP QoS: What you need to know [Electronic Version]. VoIP News*. Retrieved December 2, 2007 from http://www.voip-news.com/feature/voip-qos-service-quality-012207/.

Elliott, I. (2008). *4Q07 consumer VoIP rankings*. Retrieved March 9, 2008 from http://ikeelliott.typepad.com/telecosm/2008/02/4q07-consumer-v.html

Erlanger, L. (2005). Is VoIP ripe for attack? *InfoWorld, 27*(42), 31.

Frommer, D. (2006). *Vonage dominates home VoIP market.* Retrieved March 9, 2008 from http://www.forbes.com/2006/03/14/voip-vonage-comcast-0314markets15.html

Galbraith, R., Turnbull, J., Dearnaley, R., Hart, J., & Nolde, K. (2005). Do we need quality of service in the customer environment? *BT Technology Journal, 23*(2), 146.

Garbin, D. A., & Gharakhanian, A. (2006). Voice quality in enterprise VOIP Systems. *Business Communications Review, 36*(2), 42.

Garretson, C. (2005). VoIP security threats: Fact or fiction? *Network World, 22*(29), 14.

Garretson, C. (2007). Hype vs. reality in VoIP security. *Network World, 24*(5), 20.

Hunt, G., & Arden, P. (2005). QoS requirements for a voice-over-IP PSTN. *BT Technology Journal, 23*(2), 37.

Huston, G. (2007). Quality of service-fact or fiction? Retrieved November 27, 2007, from http://www.cisco.com/web/about/ac123/ac147/ac174/ac197/about_cisco_ipj_archive_article09186a00800c8314.html

International Telecommunications Union. (1996). Methods for subjective determination of transmission quality [Electronic Version]. Retrieved October 29, 2007 from http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.800-199608-I!!PDF-E&type=items.

Karol, M., Krishnan, P., & Li, J. J. (2005). VoIP protection and performance Improvement. *Telecommunication Systems, 28*(3-4), 351.

King, R. (2007). Line goes dead for a Web caller. *Business Week Online*, 3-3.

Kong, X., & Mase, K. (2003). Dynamic routing with endpoint admission control for VoIP networks. *IEEE International Conference on Communications, 2003. ICC '03, 3*, 1728-1732.

Kretkowski, P. D. (2006). *Top 10 VoIP predictions for 2007.* Retrieved October 5, 2007, from http://www.voip-news.com/feature/top-ten-voip-predictions-122706/

Laurence, M. (2005). How echo affects VOIP. *Business Communications Review, 35*(8), 50.

Leslie, C. (2007). Courts rebuff most of Vonage's patent arguments. *USA Today*, Retrieved November 10, 2007 from Academic Search Premier EBSCO database.

Mark, G. (2005). Connecting existing phones to VoIP. *Network World, 22*(40), 50.

Mier, E. E. (2007). SIP: A mid-course assessment. *Business Communications Review, 37*(3), 28.

Morrissey, P. (2005). Skype: Unsafe in any enterprise. *Network Computing, 16*(23), 22.

Rash, W. (2005). Can only you hear me now? Technologies That Can Secure Voice over IP. *Federal Computer Week, 19*(34), 29.

Rubino, G., Varela, M., & Bonnin, J.-M. (2006). Controlling multimedia QoS in the future home network using the PSQA metric. *The Computer Journal, 49*(2), 137.
Schulzrinne, H. G., & Rosenberg, J. D. (1998). The Session Initiation Protocol: Providing advanced telephony services across the Internet. *Bell Labs Technical Journal, 3*(4), 144.

Schulzrinne, H. G. (1999). Service for Telecom Version 2. *IEEE Internet Computing, 3*(3), 40-43.

Scoblete, G. (2007). Competitors swarm to stranded SunRocket subscribers. *TWICE: This Week in Consumer Electronics, 22*(16), 41-41.

Skype. (2007). *Skype Technical FAQ.* Retrieved October 21, 2007 from http://www.skype.com/help/faq/technical.html

Thermos, P. (2006). *Two attacks against VoIP*. Retrieved October 21, 2007 from http://www.securityfocus.com/infocus/1862/2

Varshney, U., Snow, A., McGivern, M., & Howard, C. (2002). Voice over IP. *Communications of the ACM, 45*(1), 89-96.

ViaTalk. (2007). *How to set up your ViaTalk broadband phone*. Retrieved November 27, 2007, from http://support.viatalk.com

voip-info.org. (2007). ATA. Retrieved October 22, 2007, from http://www.voip-info.org/wiki-ATA

VoIP Mechanic. (2006). How to correct VoIP echo. Retrieved November 26, 2007, from http://www.voipmechanic.com/phoneandvoicequality.htm

Waclawsky, J. G. (2007). SIP's future: Complicated and competitive. *Business Communications Review, 37*(2), 22.

Wilson, C. (2005). VoIP quality questions stir echoes. *Telephony, 246*(12), 10.

Wilson, C. (2007). Testing VoIP for quality. *Telephony, 248*(5), 33.

Yoo, C. S. (2006). Network neutrality and the economics of congestion. *Georgetown Law Journal, 94*(6), 1847.

KEY TERMS

PSTN: Public Switched Telephone Network. The traditional circuit-switch telephone network operated by local and long distance telephone companies.

SIP: Session Initiation Protocol. One of several protocols that enables VoIP communications by assisting with determining the location of the destination user, and negotiating the encoding scheme and calling features to be utilized during the VoIP conversation (viop-info.org, 2007).

Skype: VoIP service provider owned by e-Bay that uses proprietary protocols for VoIP communications with other Skype members and to the PSTN. In general, Skype phones and adapters are not compatible with SIP-based services (Waclawsky, 2007).

VoIP: Voice over Internet Protocol. Using Internet protocols for voice communications over local and wide area networks including the public Internet.

Chapter XXXIX
Emerging Security Issues in VANETs for E-Business

S. S. Manvi
REVA Institute of Technology and Management, India

M. S. Kakkasageri
Basaveshwar Engineering College, India

ABSTRACT

This chapter presents the emerging security issues in Vehicular Ad hoc Networks (VANETs) for e-business along with some of the solutions provided by the research community. The VANET will facilitate new applications for e-business that will revolutionize the driving experience, providing everything from instant, localized traffic updates to warning signals when the vehicle ahead abruptly brakes. In the emerging global economy, e-business has increasingly become a necessary component of business strategy and a strong catalyst for economic development. In near future, vehicles may be equipped with short-range radios capable of communicating with other vehicles and highway infrastructure using a VANET. However, providing security in VANETs for e-business raises privacy concerns that must be considered. The deployment of VANETs for e-business is rapidly approaching, and their success and safety will depend on viable security solutions acceptable to consumers, manufacturers and governments.

INTRODUCTION

VANET is a form of Mobile Ad-hoc NETwork (MANET) that provides communications between vehicles and external network infrastructure. VANETs are expected to have great potential to improve both traffic safety and comfort in the future (Murat, 2005; Holgar, 2005; Sascha, 2006). E-business can be conducted over VANET to facilitate business activities among users traveling in the vehicles. It is a process that relies on an automated information system. E-business methods enable companies to link their internal and external data processing systems more efficiently and flexibly, to work more closely with suppliers and partners, and to better satisfy the needs and expectations of their customers.

VANET has become a promising field of research since the world is advancing towards the vision of Intelligent Transportation Systems (Sascha, 2004; Manvi, 2006; Manvi 2007).

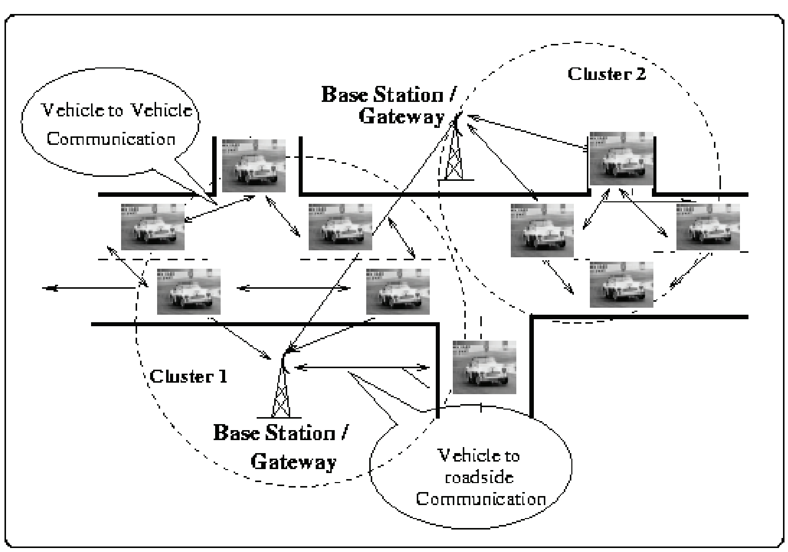
Vehicles (or nodes) in VANET are assumed to be equipped with the following.

- GPS (Global Positioning System) receiver enabling the vehicle to track its own location.
- Onboard computing devices allowing the vehicle to perform simple calculations including encryption and other vehicle's positions.
- Communication devices (Direct Short Range Communication compliant) to propagate/receive information.
- Equipment enabling to verify neighbor's position and identify obstacles.
- A set of sensors reporting crashes, engine statistics, weather conditions, etc.
- Pre-stored digital maps.
- Dedicated and secured memory.
- Its own clock to obtain an accurate timestamp.

VANET enables communications between nearby vehicles (V2V communications) and the roadside infrastructure (V2I communications). While using mostly V2V communications, VANET does not entirely rely on a fixed infrastructure, but can harness it for improved performance and functionality when it is available. A typical VANET scenario is as shown in figure 1. Vehicle to vehicle and vehicle to roadside base station/gateway communication is required for providing safety and other information services to vehicle users. Group of vehicles together may form a cluster to disseminate information among themselves as well as to other clusters and base stations.

Safety applications require reliable delivery of emergency warning messages with high reliability and low latency constraints to nearby and approaching vehicles. However, existing IEEE 802.11x standards fall short of what is required for high-speed VANET applications. In order to provide latency minimization, message prioritization, and elements of security (authorization and anonymity), in 1999, the Federal Communication Commission has allocated 75 MHz of spectrum in the 5.9 GHz band for Dedicated Short Range Communications (DSRC or 802.11p) for VANET communications. In DSRC, 5.9 GHz (5.850-5.925)

Figure 1. A typical VANET scenario



band is divided into seven 10 MHz channels, with explicit allocation of one control and six service channels to accommodate diversity of applications envisioned for VANET (DSRC, 2004; DSRC, 2005).

Safety applications that have a potential of influencing life-and-death decisions will be transmitted over the "Critical Safety of Life" channel and, thus, will have the highest priority (imminent collision warning, forward obstacle detection and avoidance, emergency message dissemination). Traffic congestion advisories and cooperative driving applications communicate over the "Public Safety Channel". Beacon signals, the backbone of VANET communication, will be exchanged on a dedicated "Control Channel", whereas private advertisements (e.g. food chain advertisements, toll collection services, etc.) will be sent on "Service channels".

Even though VANETs share similarities with MANETs like short radio transmission range, low bandwidth, low storage capacity, several new issues arise because of the unique characteristics of the vehicular context. The fundamental characteristics that differentiate VANETs from other networks are geographically constrained topology, partitioning and large-scale, vehicle mobility, Self-organization, unpredictability, power consumption, node reliability, channel capacity, and vehicle density. Some of the VANET issues related to e-business are media access control (MAC), mobility management, data aggregation, data validation, data dissemination, network congestion, privacy, and security.

Some of the important general applications of VANETs are message and file delivery, Internet connectivity, communication-based longitudinal control, co-operative assistance systems, safety services, traffic monitoring and management services, and e-business.

The chapter is organized as follows. Firstly, e-business activities over VANETs are discussed. Secondly, the security issues for e-business activities such as authentication, trust, confidentiality, non-repudiation, management of illegitimate access to public networks and detection of malicious attackers are described with some of the ongoing research works. Finally, the chapter is summarized along with some of the key terms.

E-BUSINESS OVER VANETs

In e-business, Information and Communication Technology (ICT) is used to enhance one's business. It includes any process that a business organization conducts over a computer-mediated network. A more comprehensive definition of e-business is: "The transformation of an organization's processes to deliver additional customer value through the application of technologies, philosophies and computing paradigm of the new economy."

VANET offers plethora of opportunities for providing e-business services to moving vehicles by taking advantage of short-range and inter vehicle wireless communication. E-business services may include on-demand information access either from the neighboring vehicles or the Internet, buying and selling goods and services, performing auctions, advertising, gaming, location aware services for entertainment and other activities, traffic congestion advisories, stolen vehicle tracking, software update, map download/update, traveler and tourist information, asset tracking, road hazard detection, slot booking, fleet management, accident notification, transport on demand, automated toll and parking services and time sensitive information (like weather conditions, accidents, traffic flow, traffic jams, road conditions and area maps, etc.).

Some e-business applications are already making the use of vehicular communication without considering the security aspects. For example, insurance companies install black boxes in vehicles to collect their usage data (e.g., traveled distance) and to calculate insurance costs accordingly. Another related application is vehicle tracking using GPS.

SECURITY IN VANETs FOR E-BUSINESS

Authentication

Authentication is a key element in a security system of VANETs for e-business. Due to the large number of independent network members (i.e., they do not belong to the same organization) and the presence of the human factor, it is highly probable that misbehavior will arise.

Authentication ensures that the sender of a message is correctly identified. With ID authentication, the receiver is able to verify a unique ID of the sender. The ID could be the license plate or a chassis number of the vehicle. Yet, in many cases, the actual identity of nodes does not play an important role – receivers are satisfied if they are able to verify that the sender has a certain property. For e-business applications using location information, location authentication allows to verify that the sender is actually at the claimed position.

Customer authentication in VANETs for e-business provides the following:

- Authenticity and integrity of received asynchronous information.
- Authorization of a customer for certain attributes or rights.
- Confidentiality of business critical information.
- Communication channel secrecy.
- Prevention of replay attacks.

A good authentication mechanism in VANETs for e-business eliminates the problem of hacking. Hacking is broadly defined as the unauthorized use of nodes and network resources. Hacking sites and distributing computer viruses, which cripple programs or destroy data, are serious issues. Any business that has ambitions to become a significant player in VANETs for e-business must plan, implement and maintain an appropriate security system to protect its investment. Key elements in such a system should be appropriate back-up data, a firewall system and possible mirror sites that can be activated if one site is either damaged or failed.

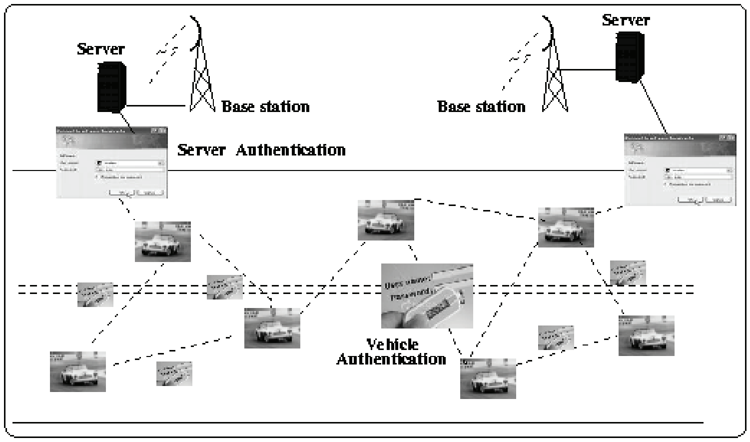
Traditional authentication methods in e-business like passwords and personal identification numbers (PINs), digital certificates using a public key infrastructure (PKI), Microchip-based devices such as smart cards or other types of tokens, database comparisons (e.g., fraud-screening applications), and biometric identifiers may be used for VANETs to provide e-business. But these methods must consider the unique characteristics of VANETs like device speed, limited battery life, smaller screen sizes, different data entry formats, and limited capabilities to transfer stored records. Hence there is a strong need for good authentication mechanisms in VANETs for e-business.

Authentication scenario for e-business in a VANET is as shown in the figure 2. In order to establish the authentication in vehicle-to-vehicle communication password mechanism can be used. For communicating vehicles with roadside infrastructure such as base station, server authentication mechanisms can be used.

Bryan (2005) proposes a set of security primitives that can be used as the building blocks of secure applications. This work analyzes the security challenges specific to VANETs and vehicular properties that can support secure systems. A set of primitives for secure applications is also proposed. Two security techniques, entanglement and reanonymizers that leverage unique vehicular properties discussed in this work are as follows.

A vehicle's relative location is defined by its entanglement with other vehicles. Each vehicle regularly broadcasts its identity (a public key) along with its signature of a current timestamp. When a vehicle receives such a broadcast, it signs the other vehicle's ID and rebroadcasts it. Vehicles traveling in opposite directions allow both streams

Figure 2. VANET authentication scenario



of traffic to perform relative localization.

An anonymization resolves some of the tension between authentication and privacy. It would rely on the observation that for almost all of the applications, a vehicle does not need to authenticate the exact identity of the other vehicle sending the information, but only the connection between the information sent and a vehicle present on the road. Drivers could use their permanent identity to authenticate to an anonymization service. The service would then provide the driver with a temporary identification that cannot be traced back to the driver. This primitive prevents spoofed identities, while still preserving driver's privacy expectations.

Ahren (2007) explained that VANET safety applications require authentication of the physical properties of the sender for security, not just traditional cryptography based identity authentication. Beacon-based Convoy Member Authentication (BCMA) is proposed in this work. BCMA relies on continued presence to determine if a vehicle is indeed in the vicinity for an extended time period. Continued presence is defined through the use of a required number of beacons before a vehicle is accepted as part of the convoy (i.e., a vehicle traveling with the message recipient). In BCMA, a vehicle only considers another vehicle as part of the convoy after it hears a threshold *'t'* beacons during *'T · (t+x)'* seconds, where *'T' is* the minimum time between Cooperative Collision Warning (CCW) beacons and *'x'* is the maximum number of acceptable lost beacons. The assumption here is that a vehicle traveling in the opposite direction or a stationary radio will be in radio range for a shorter period of time when compared to vehicles traveling in the same direction.

Chesher (2000) presents a set of simple techniques for key establishment over a radio link in peer-to-peer wireless networks. This approach is based on the Diffie-Hellman key agreement protocol, which is known to be vulnerable to the "man-in-the-middle" attack if the two users involved in the protocol do not share any authenticated information about each other (e.g., public keys, certificates, passwords, shared keys, etc.) prior to the protocol execution.

Pandurang (2006) presents a security framework for VANETs using identity-based cryptography, to provide authentication, confidentiality, non-repudiation and message integrity. Additionally it provides scalable security and privacy using short-lived, authenticated and unforgivable pseudonyms. Each vehicle and base-station has a unique identifier. These identifiers can be certified at regular periods by a Trusted Arbiter (TA). If any certificate is revoked, the TA notifies all the base-stations in the system, so base-stations have to only store Certificate Revocation List (CRL) entries that are less than a year old. Vehicles never have to download any CRLs, which provide huge savings in communication costs. The TA conducts the setup phase of the identity-based cryptosystem and computes the relevant system parameters and the master secret 'S'. Both of these are then distributed to all the base-stations in the system. The TA also generates a random secret key 'KI' for each base-station 'I' and distributes it to that base-station. The TA keeps a copy of this key in its database to help in future arbitration proceedings. The TA provides each vehicle with its unique vehicle identifier, public key certificate certifying this identifier and including a public and private key pair generated using classical algorithms like RSA. Additionally each vehicle is provided with all the public system parameters of the identity-based cryptosystem.

Base-stations have up-to-date CRLs and that they will only issue a new pseudonym only if the vehicle's credentials have not been revoked. The pseudonym includes a time-stamp indicating the last time validation of vehicle credentials by an infrastructure. Each vehicle could set its trust threshold as per the user's choice, in deciding how old pseudonyms they want to trust. Once that choice is made, validate the identity-based signature on the message to verify that the vehicle using the pseudonym actually has the private key corresponding to it. The private key could only have been generated by a base-station (or the TA) who has the master secret 'S'.

Trust

In e-business process of a VANET system, the received information has to be verified within a short time to be able to use the information as soon as possible. Trust is a critical aspect of e-business. Online purchase renders a customer vulnerable in many ways due to the lack of proven guarantees that an e-vendor will not behave opportunistically. The Internet is a complex social environment, which still lacks security. When a social environment cannot be protected through rules and regulations, people adopt trust as a central social complexity reduction strategy. Therefore, online customers have to trust an e-vendor from which they purchase; otherwise, the social complexity will avoid them involving in e-business. Trust is the key to the success of e-business in VANETs. The continuously changing connectivity to different neighbors and the not guaranteed access to an Internet gateway node make the use of certificates a challenge.

There are a variety of technological solutions available to address the issues of online trust. Most of them employ data encryption technologies, digital certificates and digital watermarks to address these issues. Trust services can be offered through a hierarchy of services such as providing public key infrastructure, certification agencies and authenticating digital certificates. Recent initiatives by VISA, Master card, Netscape, IBM, Cybercash and Microsoft in several combinations have resulted in developing new technology-based solutions such as Secure Electronic Payment protocol (SEPP), Secure Transaction Technology (STT) and Secure Electronic Transaction (SET) (Kevin, 2007). These new technologies ensure securing online financial transactions and thereby improving online trust. Trust establishment among vehicle to vehicle can be done with the help of mechanisms like public keys and digital certificates. PKI is used establish the trust between the base stations. A central certificate authority as a third party may be employed to issue the certificate to vehicles to participate in the e-business activities over the VANETs.

The summarization of the issues related to the online trust of e-business in VANETs is given in Table 1.

Anne (2007) proposes a simple, scalable and robust scheme for the distribution of revocation information in mobile ad hoc networks (MANETs) called as the MRL scheme. A MRL is a revocation list for a specific MANET instantiation. The scheme is designed for revocation of keys used to protect the ad hoc network service, and the MRLs are distributed with the routing messages. MRLs are established with the aid of trusted gateways reporting MANET nodes to a central trusted entity. The trusted gateways must be able to detect which nodes are in the MANET. This comes intrinsically with proactive ad hoc routing protocols, and may be achieved at some additional cost with reactive protocols. The scheme is intended for ad hoc networks with a planned origin, and where a common point of trust exists.

Yingying (2007) proposed a method for both detecting spoofing attacks, as well as locating the positions of adversaries performing the attacks. The method operates in the following manner. 1) An attack detector for wireless spoofing that utilizes K-means cluster analysis is designed. 2) Integrate attack detector into a real-time indoor localization system, which is also capable of localizing the positions of the attackers. 3) The positions of the attackers can be localized using either area-based or point-based localization algorithms with the same relative errors as in the normal case. This scheme is evaluated through experimentation using both an 802.11 (WiFi) network as well as an 802.15.4 (ZigBee) network. Results show that it is possible to detect wireless spoofing with both a high detection rate and a low false positive rate, thereby providing strong evidence of the effectiveness of the K-means spoofing detector as well as the attack localizer.

Sattid (2002) explained few steps to enable corporate clients to update their web content more securely. The work is on "Secure data Center" that offers security to host web sites delivering e-commerce trading, which is most cost effective and more secure. This model addresses the (Internet service provider) ISP's snags, system weaknesses and fulfills the e-business needs. The model data center is unique in design; where remote access, security, intruder detection system and other state-of-the- art equipment is in place.

Confidentiality

E-business applications require confidentiality of data exchanged between the sender and the intended receiver, e.g. bidding price of a product in online shopping. Confidentiality ensures that information content is never revealed to entities that are not authorized to receive it. Thus, a mechanism that has this attribute ensures that transmitted information can only be interpreted by the intended receivers. Secrecy and privacy are other terms that are considered synonymous with confidentiality. In the case of ad hoc networks, confidentiality is very important for protecting the transmission of sensitive information. This is especially vital considering the fact that the wireless links are easily susceptible to eavesdropping. However, this attack can be prevented by using an encryption scheme at the link level to protect the transmitted data. Of course, this

Table 1. Issues of trust aspects in VANETs for e-business

Issues	Trust aspects in VANETs for e-business
Eavesdropping	Information remains intact, but its privacy is compromised. For example, someone could get to know vehicle identity number, intercept classified information or record a sensitive conversation.
Impersonation	Information passes to a vehicle (customer) that poses as the intended recipient. Impersonation can take two forms: **1)** **Misrepresentation:** A vehicle can misrepresent itself. For example, a vehicle (customer) can pretend to be an online seller of entertainment services (like special collections of movie, songs etc.,) and that takes credit-card payments but never gives any services. 2) Spoofing**:** Pretending to be someone else. For example, a vehicle (customer) can identify itself as Internet service provider when it is not.
Tampering	Information in transit is altered or replaced and then sent to the recipient. For example, someone could change the vehicle (customer) identity or alter an order for goods.

requires efficient key distribution strategies so that keys for encrypting the transmitted traffic can be transmitted to all nodes.

Confidentiality is usually associated with encryption of data only, however there are other aspects to be considered for the case of e-business in VANETs. In e-business process the confidential data is stored in roadside units (like base stations, gateways etc.,). Access to their interfaces must be carefully controlled to allow access only to appropriate users to these highly confidential data without that data being compromised or revealed. Confidentiality also extends to the privacy requirements of the actual users and resources. In order to manage the unauthorized access of the confidential data stored in the server of the base station, encryption techniques may be employed. Encryption and decryption techniques can also be used to communicate the confidential data among vehicles.

Stephan (2007) addresses the security requirements existing in a Vehicular Network and present how to set up and implement the security architecture in a vehicular network node. The in-vehicle system architecture can be grouped into three component blocks: the main system stack (interfaces, communication, services, Human-Machine Interface (HMI)), the security and privacy components (service security, communication security and system security) and the support components (e.g. in-vehicle sensors).

Frank (2006) suggests a new approach that allows analyzing a large set of applications in vehicular communication, selecting typical representatives that will cover the requirements of a whole cluster of applications and develop a security solution for this subset. The disadvantage of this approach is the extreme complexity due to number of application characterizations, which makes it impossible to design security mechanisms for each individually. The approach operates in the following phases. 1) Create Application List, 2) Find Application Characteristics.

Application list contains all applications that might be used in vehicular communication scenarios together with a short description of an application. After the properties are defined, each application needs to be classified in every property. A meaningful class for each property is given. Importance of the property for an application, where '0' stands for irrelevant, '1' for important and '2' for very important. This is necessary so the cluster analysis algorithm can determine numerical distances between properties and applications.

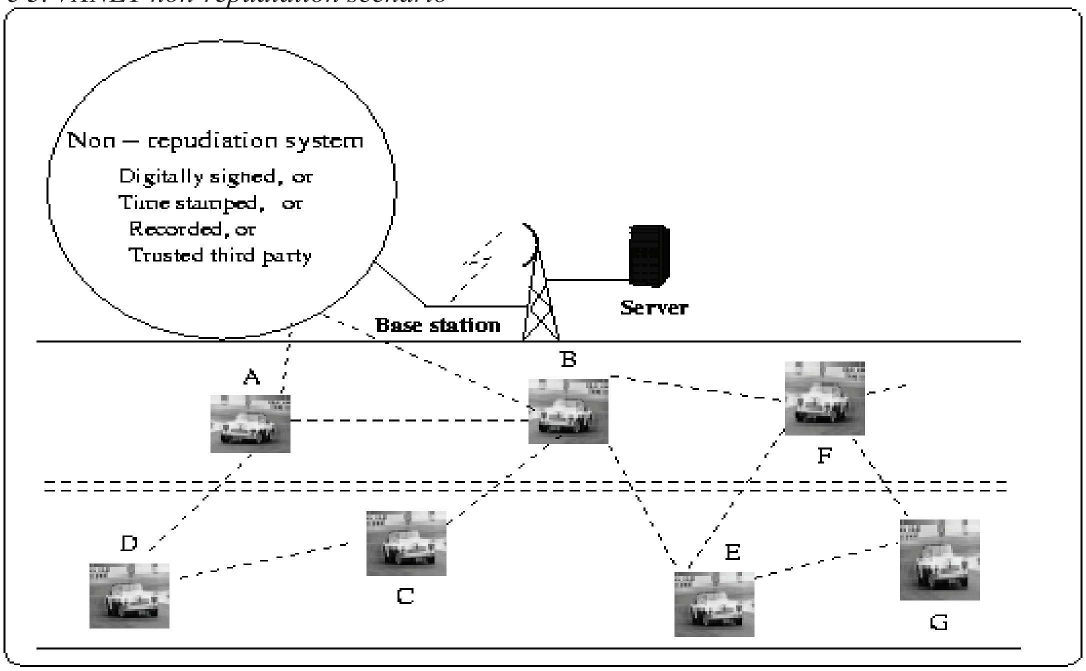
Non-Repudiation

Non-repudiation in VANETs is the concept of ensuring that a vehicle in a dispute cannot repudiate, or refute the validity of a statement or contract. Non-repudiation ensures that a transferred information has been sent and received by the parties claiming to have sent and received the information. For example, drivers causing accidents should be reliably identified. Non-repudiation can be obtained through the use of digital signatures (function as a unique identifier for an individual, much like a written signature), confirmation services (like the message transfer agent can create digital receipts to indicated that information is sent and/or received), and timestamps (timestamps contain the date and time a document was composed and proves that a document existed at a certain time).

Non-repudiation model in VANETs must have the ability of a vehicle involving in business to prove or disprove having previously sent or received a particular business message to or from another vehicle involving in business. Consider the following example as shown in figure 3.

Vehicle involving in business A has agreed to download an audio/video song from another vehicle B. In the course of this agreement, vehicle A has sent a business message to vehicle B agreeing to download the audio/video song at a set price. Later, though, vehicle A disputes the

Figure 3. VANET non-repudiation scenario



original price and denies having sent a message in which they agreed to pay that price.

If a reliable non-repudiation system has been in place, vehicle B can disprove vehicle A's claim by producing a document from vehicle A specifying the amount vehicle A agreed to pay. Further, if this original document is digitally signed, time stamped, recorded, and secured by a trusted third-party source, the validity of this document has full legal recourse. Non-repudiation, or the ability to provide legal evidence of the involvement of a denying party, is a requirement for critical business messages.

Sumair (2007) proposed security scheme to meet the requirements like authentication, data consistency, non-repudiation, privacy, and real-time constraints. In this scheme non-repudiation is made possible to reliably identify vehicles from the AutoCore (AUTOmated COllision REporting) messages they transmit such that drivers cannot deny their vehicles were the source of these messages, allowing investigators to determine liability in the event of a collision. Non-repudiation is achieved as follows: (1) a vehicle cannot claim to be a different vehicle, because it signs messages with its own private keys. Furthermore, ELPs (Electronic License Plates) are unique and only one vehicle holds the corresponding ELPPr (ELP of a signing key pair) in its TPD (Tamper-Proof Device). (2) a vehicle cannot deny having sent messages because a message is signed with AnonCredPr, (Anonymous Credentials key pairs) which belongs to the vehicle and was generated by the vehicle in the first place. Timestamps included in each message guard against message replay attacks.

Christine (2007) proposed a Wireless Access in Vehicular Environments (WAVE) based protocol for the secure and anonymous propagation of vehicle safety broadcast messages. A trusted device S transmitting a broadcast message M signs it with its private key S−1 and transmits: {M, S, {M} S−1}. The recipient verifies the signature using the unsigned message, S's public key and the signature. If the signature is verified and the public key certificate is deemed valid, then the message is processed. Otherwise, it is discarded. In this way non-repudiation is achieved.

Managing Illegitimate Access of Public Access Networks

The proliferation of 802.11-based access points, base stations in VANETs, hotspots and their ad hoc extension as meshes has created a demand for the ability to securely mutually authenticate the access point and the mobile user. IEEE 802.1X defines a mechanism for authenticating the client and access point, and controlling access to the wireless 'port'. It requires a pre-shared secret key between the user and network, and as such is most appropriate for the e-business in VANET environment. For e-business in VANETs the access points, base stations in VANETs, hotspots often have business models that need to support single use authorization such as prepaid cards, as well as allow an access point and its corresponding scarce RF spectrum to be utilized by more than one service provider. Typically these public access networks utilize a web-based front-end to an authentication system that performs packet filtering on some combination of the MAC and IP address, or proprietary client software that reduces interoperability options. Because of the ease of spoofing both the MAC and IP addresses, it is possible to deny legitimate users access to the network, or use their credentials for unauthorized access to the network.

Using firewalls and some encryption techniques can prevent illegitimate access of public access networks in VANETs. To manage the hotspot's access point, web based authentication and IEEE 802.11x mechanisms can be used. Firewalls are used to provide the Internet connection to the base stations.

Jong (2005) explained the balance between privacy and audit requirements in VANETs. This explains the feasibility study on the topic of using symmetric cryptosystems to build VANETs with balanced privacy and auditability, but with lighter requirements on communication and computation.

Maxim (2005) explored the approach of secure message aggregation for VANETs, the long-time trademark of resource constrained sensor networks. Instead of letting the de facto flooding approach take care of message dissemination in a VANET, this is delegated only to selected vehicles that share a similar view of their environment.

Krishna (2005) dealt the problem of providing location privacy in VANET by allowing vehicles to prevent tracking of their broadcast communication. This work also addresses the problem of allowing any vehicle to be able to achieve unlinkability between two or more of its locations in the presence of tracking by an adversary. Three typical classes of VANET applications, cooperative driving, probe vehicle data, and location-based service are considered. In the cooperative driving application, adequate equipped vehicles maintain a very short separation (intra-convoy spacing) between each other and move smoothly with the same pre-defined speed (convoy speed). These vehicles communicate with each other frequently either directly or via communication equipments on roadside. Vehicles broadcast their status information (e.g. speed, location, acceleration) every 500 ms.

Malicious Attackers

Malicious attackers deliberately attempt to cause harm via the e-business applications available on the VANET. In many cases, these attackers will have specific targets, and they will have access to more resources. Two types of misbehavior mainly cause the intentional non-cooperative behavior of a participating node in VANETs for e-business. They are classifieds into two types: 1) selfish behavior, e.g., nodes that want to save power, CPU cycles, and memory, 2) malicious behavior which is not primarily concerned with power or any other savings but interested in attacking and damaging the network. When the misbehavior of a node manifests as selfishness, the system can still cope with it since this misbehavior can always be predicted. A selfish node will always behave in a way that maximizes its benefits, and as such, incentive can be used to ensure that cooperation is always the most beneficial option. However, when the misbehavior manifests as maliciousness, it is hard for the e-business system to cope with it, since a **malicious** node always attempts to maximize the damage caused to the system, even at the cost of its own benefit. As such, the only method of dealing with such a node is detection and ejection from the network.

Node misbehavior due to selfish or malicious reasons or faulty nodes can significantly degrade the performance of VANETs. Malicious node misbehavior of a VANET may involve in packet forwarding. It can be generally divided into two types of misbehavior, forwarding and routing. Some common forwarding misbehavior are packet dropping, modification, fabrication, timing attacks, and silent route change. Packet dropping, modification, and fabrication are self-explanatory. Timing misbehavior is an attack in which a malicious node delays packet forwarding to ensure that packets expire their Time-To-Live (TTL), so that it is not immediately obvious that it is causing the problem. A silent route change is an attack in which a malicious node forwards a packet through a different route than it was intended to go through. A malicious node with routing misbehavior attacks during the route discovery phase.

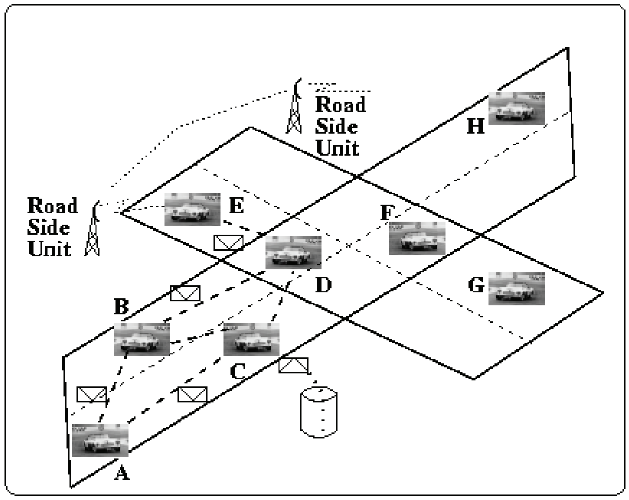
Three common attacks of this type are black hole, gray hole, and worm-hole. A black-hole attack is one in which a malicious node claims to have the shortest path and then when asked to forward drops the received packets. In a gray-hole attack, which is a variation of the black-hole attack, the malicious node selectively drops some packets. A worm-hole attack, also known as tunneling, is an attack in which the malicious node sends packets from one part of the network to another part of the network, where they are replayed. In general the selfish behavior of a node can be generally classified as either self-exclusion or non-forwarding. The self-exclusion misbehavior is one in which a selfish node does not participate when a route discovery protocol is executed. This ensures that the node is excluded from the routing list of other nodes. This benefits a selfish node by helping it save its power, as it is not required to forward packets for other nodes.

A reputation model is an effective way to prevent the intentions of such selfish nodes. Since a node does not forward packets for other nodes in the networks, it is denied any cooperation by other nodes. So, it is in the best interest of a selfish node to be cooperative. On the other hand, the non-forwarding misbehavior is one in which a selfish node fully participates in route discovery phase but refuses to forward the packets for other nodes at a later time. This selfish behavior of a node is functionally indistinguishable from a malicious packet dropping attack. Since reputation-based systems can cope with any kind of observable misbehavior they are useful in protecting a system. Reputation and trust-based systems enable nodes to make informed decisions on prospective transaction partners.

Integrity ensures that system assets and transmitted information are modified only by authorized parties. Modification includes writing, changing, changing status, deleting, creating, and the delaying or replaying of transmitted messages. It ensures that data is not altered in an unauthorized manner during transmission. This alteration could be due either to accidental factors such as the vagaries of the wireless links or malicious attackers. An adversary could manipulate data by insertion, deletion, or substitution of data.

Consider the following example scenario as depicted in figure 4 for the malicious attackers in VANETs. Vehicle A is the source and vehicle E is the destination. Vehicles B, C and D are the intermediate vehicles, which are involved in packet forwarding. For every packet, a vehicle keeps track of the behavior of the next-hop

Figure 4. Misbehavior scenario



node and remembers whether it has forwarded the packet correctly. Assume that the vehicle C misbehaves by dropping the packet instead of forwarding. When the source vehicle A learns about the malicious status of vehicle C, it changes the packet transmission path through the intermediate vehicles B and D.

Florian (2005) gave an overview on the privacy issues in VANETs from a vehicle manufacturer's perspective and introduced an exemplary approach to overcome these issues. This approach operates in three phases. The initialization phase where the systems of a vehicle are set up. The operational phase as the major mode of operation, where vehicles can send messages signed according to a chosen pseudonym. And the credential revocation phase, where predefined situations can lead to the disclosure of a vehicle's real ID and the shutdown of its system. Markus (2005) explained vehicle-to-infrastructure communication for active safety, particularly between vehicles and traffic lights. Most of the services offered by roadside units or other cars will require a secure authentication mechanism in order to be able to trust the information and prevent tampering and unauthorized use. This authentication process must not reveal the identity of the vehicles involved.

Kevin (2007) presents an analysis framework that allows general decomposition of existing reputation systems. From this analytical framework, reputation systems can be decomposed, analyzed, and compared using a common set of metrics. This framework facilitates insights into the strengths and weaknesses of different systems and comparisons within a unified framework. Classification of attacks against reputation systems, analyzing what system components are exploited by each attack category are discussed by providing specific examples based on real systems. Characterizing existing defense mechanisms for reputation systems, discussing their applicability to different system components and their effectiveness at mitigating the identified attacks is also explained.

Position dissemination is crucial for geographic routing; forged position information has severe impact regarding both performance and security. Leinmullar (2005) proposed a detection mechanism that is capable of recognizing nodes cheating about their position in beacons (periodic position dissemination in most single-path geographic routing protocols, e.g. GPSR). This detection technique does not rely on additional hardware or special nodes, which contradicts the ad hoc approach. Instead, this mechanism uses a number of different independent sensors to quickly give an estimation of the trustworthiness of other node's position claims without using dedicated infrastructure or specialized hardware. This position verification system successfully discloses nodes disseminating false positions and thereby widely prevents attacks using position cheating.

Golle (2004) proposed a general approach to evaluating the validity of VANET data. A node searches for possible explanations for the data it has collected based on the fact that malicious nodes may be present. Explanations that are consistent with the node's model of the VANET are scored and the node accepts the data as dictated by the highest scoring explanations. Generating and scoring explanations rely on two assumptions: 1) nodes can tell "at least some" other nodes apart from one another and 2) a parsimony argument accurately reflects adversarial behavior in a VANET.

FUTURE TRENDS

VANET is a promising approach for future intelligent transportation system (ITS). In the future, e-business in VANETs will become a critical competitive strategy that will revolutionize the global economy. In order to predict the future security issues in VANETs, it is important to understand the application development trends in VANETs. Some common advanced applications in VANETs for e-business are as follows. 1) Online store – can facilitate online payment and credit card transactions. 2) Content management system – allow you to update your site easily. 3) High-end interactive website – can provide a rich customer experience. 4) Online training application for employees - can increase productivity. 5) Business intranet – a secure internal website to increase efficiency. All these applications can make use of some intelligent technologies like cognitive agent technology to deal with security, since they provide human kind of reasoning. However, there is a need to look at the problems in implementation of cognitive multi agent systems and development of security algorithms using agents.

SUMMARY

VANETs combine short-range communications, with the scalability and mobility of classic *ad hoc* networks, in order to support a number of e-business applications. E-business systems inherently possess a higher degree of risk than mainstream applications, and thus require a greater degree of security. Because of this risk, security should be considered as a fundamental aspect of an e-business system design in VANETs.

We presented security aspects such as authentication, trust, confidentiality, non-repudiation, management of illegitimate access to public networks and detection of malicious attackers of e-business in VANETs, along with some of the ongoing research works. To become a real technology that can guarantee public safety on the roads, VANETs need an appropriate security architecture that will protect them from different types of security attacks. In VANETs, the high-speed mobility and the large number of network entities call for time-efficient and cost-saving authorization and authentication models, as well as fast attacker detection, revocation and trace back schemes.

REFERENCES

Ahren, S., Mark, L., & Adrian, P. (2007). Efficient mechanisms to provide convoy member and vehicle sequence authentication in VANETs. In B. Crispo (Ed.), *The Third International Conference on Security and Privacy in Communication Networks.* IEEE: France.

Anne, M., Pal, S., Leif, N., & Oivind, K. (2007). On the distribution of revocation information in ad hoc networks. In P. Dini (Ed.), *The Third International Conference on Wireless and Mobile Communications* (p. 84). IEEE: USA.

Bryan, P., & Adrian, P. (2005). *Challenges in securing vehicular networks.* In N. Spring (Ed.), *The Fourth Workshop on Hot Topics in Networks* (pp. 49-55). ACM SIGCOMM.

Chesher, M., & Kaura, R. (2000). The E-business (R) evolution, Hewlett-Packard Professional Books, from htttp://www.interhack.net/pubs/network-security/

Christine, L., & Michel, B. (2007). Secure Anonymous Broadcasting in Vehicular Networks. In C. Ahlund & S. Cherkaoui (Eds.), *The First IEEE LCN Workshop On User MObility and VEhicular Networks (ON-MOVE)* (pp. 1300-1313), IEEE: Ireland.

Dedicated Short Range Communications (2005), from http://www.itsa.org/itsa/files/pdf/DSRCIC-WhitePaper.pdf

Dedicated Short Range Communications (DSRC) Service (2004). *The FCC DSRC website*, http://wireless.fcc.gov/services/its/dsrc/

Florian, D. (2005). Privacy issues in vehicular ad hoc networks. In George D. & David M. (Eds.), *The Fifth International Workshop On Privacy Enhancing Technologies* (pp. 197-209). Springer.

Frank, K., Zhendong M., & Elmar S. (2006). Security engineering for VANETs. In K. Borras (Ed.), *The Fourth Workshop on Embedded Security in Cars* (pp. 374-381). ACM: Germany.

Golle, P., Greene, D. H., & Staddon, J. (2004). Detecting and correcting malicious data in VANETs. In H. Boral & P. Larson (Eds.), *The First ACM Workshop on Vehicular Ad Hoc Networks* (pp. 29-37). ACM: USA.

Holger, F., Marc, T. M., Matthias, T., Roland, K., Hannes, H., & Wolfgang, E. (2005). Studying vehicle movements on highways and their impact on ad-hoc connectivity. In M. Satyanarayanan (Ed.), *The Eleventh Annual International Conference on Mobile Computing and Networking* (pp. 26-27). ACM: Germany.

Jong, Y. C., Markus, J., & Susanne, W. (2005). Balancing auditability and privacy in vehicular networks. In Y. Xiaos (Ed.), *The First ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks* (pp. 79-87). ACM: New York, USA.

Kevin, H., David, Z., & Cristina, R. (2007). *A Survey of attack and defense techniques for reputation systems.* From http://www.cs.purdue.edu/homes/zagedj/docs/reputation_survey.pdf

Krishna, S., Leping, H., Mingyan, L., Radha, P., Matsuura, K., & Sezaki, K. (2005). CARAVAN: Providing location privacy for VANET. In A. Amer (Ed.), *The Embedded Security in Cars* (pp. 367-383). ACM: Germany.

Leinmuller, T., Schoch, E., Kargl, F., & Maihofer, C. (2005). Influence of falsified position data on geographic ad-hoc routing. In R. Molva, G. Tsudik, & D. Westhoff (Eds.), *The second European Workshop on Security and Privacy in Ad hoc and Sensor Networks* (pp. 102-112). Springer.

Manvi, S. S., Kakkasageri, M. S., Jeremy, P., & Alex, R. (2006). Multi agent systems as a platform for VANETs. In H. Nakashima, P. Michael, Wellman, G. Weiss, P. Stone (Eds.), *Autonomous Agents and Multi Agent Systems (AAMAS)* (pp. 35-42), ACM: Japan.

Manvi, S. S., Kakkasageri, M. S., & Jeremy, P. (2007). Information search and access in vehicular ad hoc networks (VANETs): An agent based approach. In H. R. Arabnia (Ed.), *The International Conference on Communication in Computing (CIC-2007), The 2007 World Congress in Computer Science, Computer Engineering and Applied Computing* (pp. 23-29), CSREA Press: USA.

Markus, S., Florian, K., Timo, K., & Florian D. (2005). Secure communication for intersection assistance. In H. Rohling (Ed.), *The Second Inter-*

national Workshop on Intelligent Transportation (pp. 156-163). IEEE: Germany.

Maxim, R., & Jean, H. (2005). The security of vehicular ad hoc networks. In Y. W. Law (Ed.), *The Third ACM workshop on Security of ad hoc and sensor networks* (pp. 11-21). ACM: USA.

Murat, C., Martin, M., Bernd, R., & Andreas, L. (2005). Collection of dedicated information in vehicular ad hoc networks. In L. Katz & R.M. Sorrentino (Eds.), *12th World Congress on Intelligent Transport Systems* (pp. 195-207). Cognizant: U.S.A.

Pandurang, K., Arati, B., & Wade, T. (2006). An identity-based security framework For VANETs. In W. Holfelder, D. Chrysler, & D. B. Johnson, (Eds.), *The Third international workshop on Vehicular ad hoc networks* (pp. 94-95), ACM: USA.

Sascha, S., Holger, F., Matthias, T., & Wolfgang, E. (2004). Traffic view: Traffic data dissemination using car-to-car communication. *ACM Mobile Computing and Communications Review (MC2R), 8*(3), 6-19.

Sascha, S., Holger, F., Matthias, T., & Wolfgang, E. (2006). Vehicular ad-hoc networks: Single-hop broadcast is not enough. In H. Rohling (Ed.), *Third International Workshop on Intelligent Transportation* (pp. 49-54), IEEE: Germany.

Sattid, M. M., Garner, B. J., & Nagrial, M.H. (2002). Information security standards for e-businesses. In E. Callaway (Ed.), *The Eighth International Conference on Communication Systems* (pp. 641-645). IEEE: USA.

Stephan, E. (2007). A security architecture concept for Vehicular Network Nodes. In C. H. Kam, (Ed.), T*he Sixth International Conference on Information, Communications and Signal Processing*. IEEE: Singapore.

Sumair, U. R. & Urs, H. (2007). Secure Crash Reporting in Vehicular Ad hoc Networks. In B. Crispo (Ed.), *The Third International Conference on Security and Privacy in Communication Networks ((SecureComm 2007)* (pp. 39-68), IEEE: France.

Yingying, T., Wade, M., & Richard, P. (2007). Detecting and localizing wireless spoofing attacks. In R. Cruz (Ed.), *The Fourth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks* (pp. 193-202). IEEE: USA.

KEY TERMS

Authentication: Is needed in order to be sure about the identity of the sender or receiver of a message. The attack is called masquerading, that is pretending to be somebody else.

Confidentiality: Is the prevention of unauthorized disclosure of information concerns the content of a message. Only the sender and the receiver are supposed to know the content. Attacks include message interception (man-in-the-middle attacks), content release to other parties, etc.

Eavesdropping: Is a passive attack. An attacker listens to the communication and gets information about the content of the message.

E-Business: Is defined as the transformation of an organization's processes to deliver additional customer value through the application of technologies, philosophies and computing paradigm of the new economy.

Integrity: Ensures that only authorized parties modify system assets and transmitted information. Modification includes writing, changing, changing status, deleting, creating, and the delaying or replaying of transmitted messages.

Node Misbehavior: Is the intentional non-cooperative behavior of a node and is caused by two types of misbehavior: selfish behavior, e.g., nodes that want to save power, CPU cycles, and memory, and malicious behavior which is not primarily concerned with power or any other savings but interested in attacking and damaging the network.

Non-Repudiation: Ensures that a transferred information has been sent and received by the parties claiming to have sent and received the information.

PKI (Public Key Infrastructure): Is the typical security architecture used for networks where the presence of online authorities is not always guaranteed.

Trust: Is based on a combination of judgment or opinion based on face-to-face meetings, or recommendations of colleagues, friends and business partners.

Vehicular Ad hoc Network (VANET): Is a form of ad-hoc network that enables communications between nearby vehicles (V2V communications) and road-side infrastructure (V2I communications).

Chapter XL
Telecommunications Network Planning and Operations Management in an Academic Environment:
The Case Study of the Aristotle University of Thessaloniki

Sotirios K. Goudos
Aristotle University of Thessaloniki, Greece

Angeliki Z. Agorogianni
Aristotle University of Thessaloniki, Greece

Zaharias D. Zaharis
Aristotle University of Thessaloniki, Greece

ABSTRACT

Network planning and management in a large organization such as a university, is a complex task. A large university is an inherently demanding environment in terms of telecommunications services offered and technologies used. This chapter presents and discusses network management and planning issues in the Aristotle University of Thessaloniki, Greece. Examples of network management procedures are given. We present the data model for service provision. The network expansion and cost reduction case studies are discussed. A SWOT analysis about University's Telecommunications Center is given. Finally, we discuss the migration analysis for future upgrades that will fully enable the use of emerging technologies such as Voice over IP (VoIP).

INTRODUCTION

The Aristotle University of Thessaloniki (AUTh) is the largest university in Greece. It comprises nine Faculties organized into 39 Schools, as well as three independent Schools. Some educational and administrative facilities are located off campus for practical and operational reasons. The telecommunications network of AUTh consists of more than 40 digital Private Branch eXchanges (PBXs), which provide advanced telephony services to about 6500 users. Additionally, a distance learning service for all campus staff is available through adequately equipped classrooms. The Telecommunications Center is the administrative unit in AUTh that manages the network, provides the services and plans future upgrades.

The main objective of this chapter is to provide details about network management and network planning issues in such a large organization. We highlight some periodic operation management tasks. Service provision issues are also discussed. We present a new data model for telecommunications service provision based on existing Service Oriented Architecture (SOA) models, namely the Open Group SOA (OpenGroup, 2008) reference model and the Governance Enterprise Architecture (GEA) service object model (Goudos, Loutas, Peristeras, & Tarabanis, 2007; Peristeras & Tarabanis, 2004). Examples of use cases for telecommunications services are given. Most PBX vendors today have shifted their products from Time Division Multiplexing (TDM) based to Voice over IP (VoIP) based architectures. Consequently, we present the migration analysis to VoIP architectures (Douskalis, 2000; Schulzrinne & Rosenberg, 1999).

Network planning and management are common issues among researchers. A literature review about routing management issues in voice and data networks is given in (Medhi, 2007). Network management related issues include quality management (Okuno & Okuda, 2003) and governance modes for service development (Van den Ende, 2003). Customer Relationship Management (CRM) using data mining techniques are presented in (Azevedo, Vellasco, & Passos, 2001) while a management approach of infrastructure telecommunications projects is given in (Mezher, Abdul-Malak, & Dayya, 1999).

The use of proprietary technologies over the past years has limited the telephone usage to being only a communication device. The integration of the voice infrastructure with web-based applications has enhanced the telephone's functionality and thus transformed it into an information tool. The rise of open client-server voice-web protocols, such as the Voice eXtensible Markup Language (VoiceXML), has enabled an Internet-like explosion in voice application development and delivery (Uppaluru, 2003). These new open business models that have emerged are among the factors that have to be taken into account for network planning. Network planning in public voice networks (Pravda & Vodrazka, 2007) and development of tools for new services (Atkins, 2004) are also issues found in the literature. The current trend in network operation is network convergence to reduce operational expenses and eliminate additional expenditures on multiple parallel core networks (McCormick, 2007). Companies are deploying Passive Optical Networks (PONs) to deliver broadband applications like video on demand and high-speed access to the Internet. VoIP service using Session Initiation Protocol (SIP) can also be delivered over PONs (Engebretson, 2007). Broadband networks deployment from service providers requires robust plans for providing various types, amounts, and locations of services at competitive prices (Bollapragada, Morawski, Pinzon, Richman, & Sackett, 2007). Due to their appeal to a broad base of customers, driven by economic incentives, improved productivity and creation of new services, VoIP networks may become an attractive value proposition to service providers. However, important issues like Legal Interception (LI) and emergency calls are not yet implemented by VoIP platforms. VoIP networks enable additional revenue opportunities through new services coupled with reduced network operational costs (Prabhakar, Rastogi, & Thottan, 2005).

The aforementioned issues are highly relative to modern university network management and planning. The continuous monitoring of telecommunications market changes in both technologies and costs is a common task for any university network manager. This is due to the fact that a university has in several occasions specific and demanding needs regarding telecommunications services (e.g., videoconference sessions, remote sensor networks). Therefore appropriate telecommunications technology and service provider selection plays an important role in fulfilling the aforementioned needs.

This chapter is subdivided into five sections. Section 2 provides a comprehensive description of AUTh's network and its periodic management procedures. Section 3 deals with AUTh's telecommunication service provision. The network planning, which includes the expansion procedure, cost reduction case studies, a SWOT analysis and the migration analysis, is given in section 4. Finally, the conclusions are given in section 5.

AUTH'S TELECOMMUNICATIONS INFRASTRUCTURE-NETWORK TOPOLOGY AND OPERATIONS MANAGEMENT

Network Description and Topology

AUTh's private digital network is built as star topology. The network is ordered in two layers: the internal, which covers the University Campus, and the external, which connects the university campus with off-premise facilities, as well as with university clinics and other Greek universities. Hence, all internal calls are free of charge. The university's network wiring complies with the standards of structured cabling (Semenov, Strizhakov, & Suncheley, 2002). Furthermore, the university's private digital network offers a variety of advanced features to all members of the academic community like Interactive Voice Response (IVR) applications, videoconferencing over ISDN or IP and call center activities. More details about the services offered are given in the following sections.

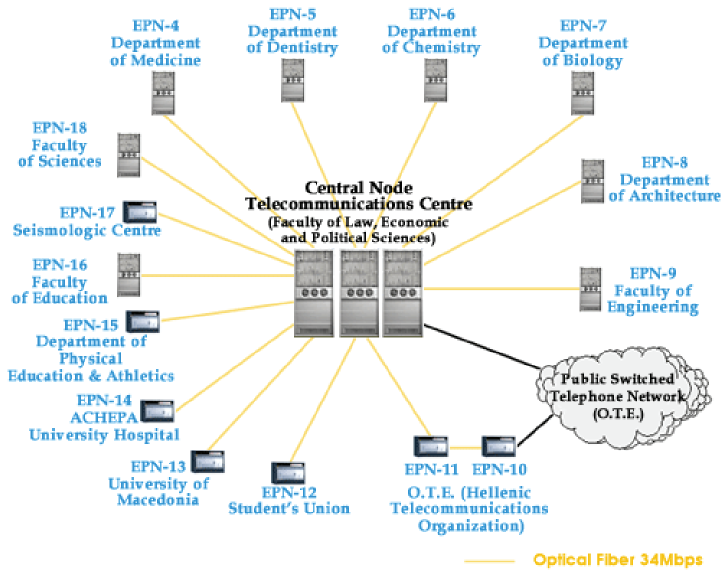
The university's Private Branch eXchange (PBX) is an AVAYA Definity G3r. In the heart of the system resides a RISC (Reduced Instruction Set Computer) processor capable of handling up to 25,000 internal users, 4000 Central Office (CO) trunks, and with 100,000 Busy Hour Call Completion (BHCC). The central node is connected through multimode fibers (34Mbps) to 17 internal nodes all over the campus. The incoming traffic is separated from the outgoing using distinct trunk lines (see Figure 1). There are 300 incoming and 300 outgoing trunk lines that connect the central node to the public telephone network. The main protocols supported by University's PBXs are Integrated Services Digital Network (ISDN), Asynchronous Transfer Mode (ATM) and TCP/IP.

Several university departments and other academic units are sited outside the university campus. Some of these units are located in the city of Thessaloniki, whilst others reside in other cities of Northern Greece. The off-premise facilities are connected to the campus voice network through ISDN Primary Rate Interface (PRI) (2.048Mbps) tie lines offering the same quality of service to all university users. A connection to voice networks of other Greek universities also exists through a VoIP (Douskalis, 2000; Schulzrinne & Rosenberg, 1999) gateway. The total number of nodes connecting to the university's voice network is 45 serving about 6500 users.

Network Management

This section presents the operations for network management. Network Management includes

Figure 1. Campus voice network topology



several daily, weekly, monthly and annual procedures. These procedures are grouped in Table 1 that comprises the AUTh's generic network management framework.

For all the aforementioned procedures two types of software tools are used. Commercial tools, available from different network equipment vendors, and custom applications, created especially for the University's network. Both types include web interfaces for easier use. The procedures related to the expansion planning and the user's requests are discussed later in this chapter. We present examples of tasks related to the network availability and applications concerning maintenance procedures.

The trunk status and the traffic monitoring are crucial daily tasks in order to ensure high network availability. There are 50 trunk groups covering AUTh's connection with other nodes and telephony companies. It is obvious that trunk status monitoring cannot be performed manually in regular intervals. Therefore a custom application has been developed for automated trunk status monitoring in an hourly basis 24 hours a day. This tool uses an automated procedure to login into the central node. Then, it issues a set of PBX commands for trunk group status and alarm monitoring. The results are captured in a text file and stored in a database. The application then publishes the results in a user-friendly manner in a web page in the Telecommunications Center intranet. This page shows only the disconnected and out-of-service lines. Alarms caused by PBX equipment malfunction are also shown. The application logic is shown in the flowchart of Figure 2. Administrators check often this web page during working hours.

In the case of disconnected lines or existing alarms, troubleshooting procedures are followed. Troubleshooting operations depend on the trunk group line type, the line speed and other factors. Troubleshooting procedures are different for every alarm type. These are documented in vendor's user guides and manuals.

Traffic monitoring is a daily task. This includes checking blockage measurements and timeslot availability among nodes for the previous day. Again, this is performed using an automated procedure. In this way, the network administrators can locate possible problems or congestion in network resources. In case of continuous congestion or high blockage percentage in certain nodes, decisions

Table 1. AUTh's generic network management framework

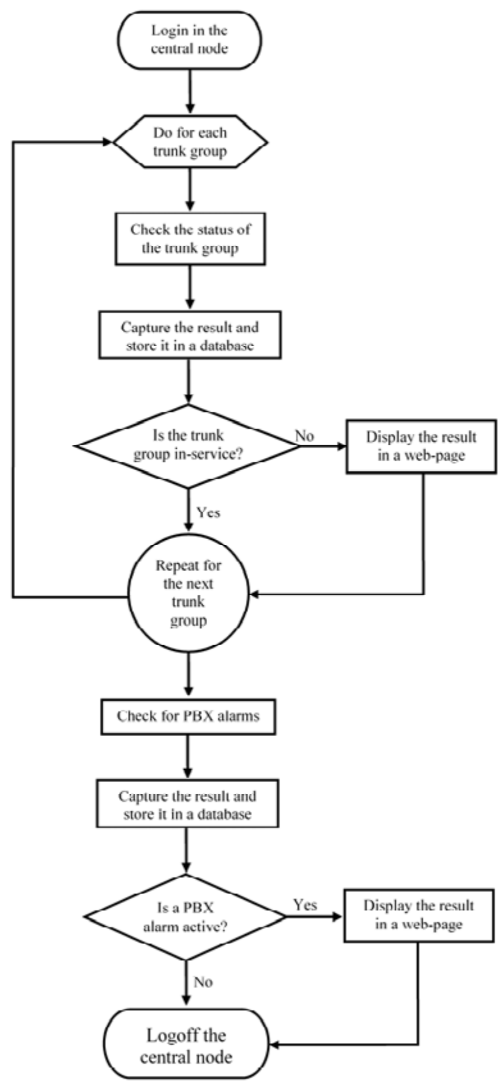
Management Task Description	Frequency			
	Daily	Weekly	Monthly	Annual
Procedures to ensure high Network Availability	x			
Service Provision Procedures	x			
Troubleshooting procedures	x			
Hardware Maintenance procedures				x
Software Maintenance procedures	x	x		
Network Documentation-Statistical Reports			x	
Expansion Planning procedures				x
Toll fraud prevention procedures	x	x	x	

for better network resources allocation are made. Network expansion planning decisions strongly depend on these daily traffic reports.

The backup procedure is very crucial for application maintenance. Therefore, backup is performed in a daily and weekly basis and concerns several databases, document files and forms used in daily service provision, executable files used in the processing of the database records, and finally all the web-pages and other supplementary files that compose the website of the Telecommunications Center. The database records and files must go through full backup in order to ensure robust administration. The backup procedure is automated and is performed in a backup server. Due to the large amount of data transferred between servers during the backup procedure the local network is congested. Therefore, the backup is performed only at night out of working hours.

Except the aforementioned procedures, sanity configuration check is also an important maintenance task. This is achieved through an automated scheduled maintenance procedure, which is performed every night. Specifically, the central node checks the status of the operating system, the system database and the storage devices hardware. If an error occurs then alarms are triggered. There are three error types: major alarms, minor alarms and warnings. Major alarms are the most severe errors that have to be resolved as soon as possible and may result to service disruption.

Figure 2. Trunk status application flowchart



Another important and necessary task for AUTh's network management is toll fraud detection and prevention. To describe toll fraud detection procedures the details of AUTh's call restrictions have to be given at first. The local area calls are free of charge. All AUTh's network users use 6-digit personal authorization codes (PACs) for making long distance and international calls. A call using a valid PAC can be made from any university's extension. The security tasks and procedures in AUTh's network aim at preventing people from using 6-digit codes without proper authorization. An authorization code violation is any attempt for long distance and international calls from a university extension using an invalid PAC. It is obvious that if a user attempts several trials the probability of finding a valid code increases. The PBX software has monitoring capability for authorization code violations. Therefore the PBX stores temporarily these attempts for an hour period. If more than three invalid attempts are made from the same extension a security violation alarm is triggered and a phone rings alerting the administration staff. Then the administration staff checks the violation details and makes a decision either to immediately disconnect this extension or to make a warning call. This decision depends on the type of violation. For example an immediate extension disconnection occurs when violations show random 6-digit patterns. It is obvious in this case that someone is trying to find a valid PAC. In case of violations that occur in weekend or out of working hours, there are automated procedures that record every unauthorized trial. Every hour in 24 basis a custom application captures the temporarily violation details from the central node and stores them in a database. Every week's records are examined in order to find similarities with existing codes or organized violations attempts (e.g., attempts using several random 6-digit numbers). If a toll fraud problem is found measures are taken. Other toll fraud detection procedures have also been used like search using data mining algorithms in Call Data Records (CDR). In (Hilas, Goudos, & Sahalos, 2006) a time series framework is proposed to classify user profiles from CDR data and detect possible toll fraud.

Concluding, we may define a set of best network management practices:

- Check alarms every day and respond quickly in case of major alarms.
- Report to service providers support service in case of disconnected CO lines or WAN links.
- Implement an automated backup procedure for all crucial files (operating systems, database records e.t.c.)
- Document in detail troubleshooting procedures.
- Create a toll fraud detection framework using CDR data.
- Check traffic load reports periodically and plan future upgrades accordingly.

Except the well-defined technical procedures the human factor remains always one of the most important factors in network management. A study of failures (Kuhn, 1997) in the US Public Switched Telephone Network has shown that human intervention plays a key role to large networks reliability. Therefore human resources management is very important. Telecommunication Center acknowledges this factor and for this reason recruiting procedures (Lee, 2005a, , 2005b) for network administrators are very strict and demanding. Staff members attend frequently educational programs and seminars. Any further analysis of human resource management issues is out of the scope of this chapter.

Service Provision – Administrative Procedures

In this section we outline the main services offered to the academic staff. The data model for this service provision is also described. The billing policy for both academic and non-academic staff is presented.

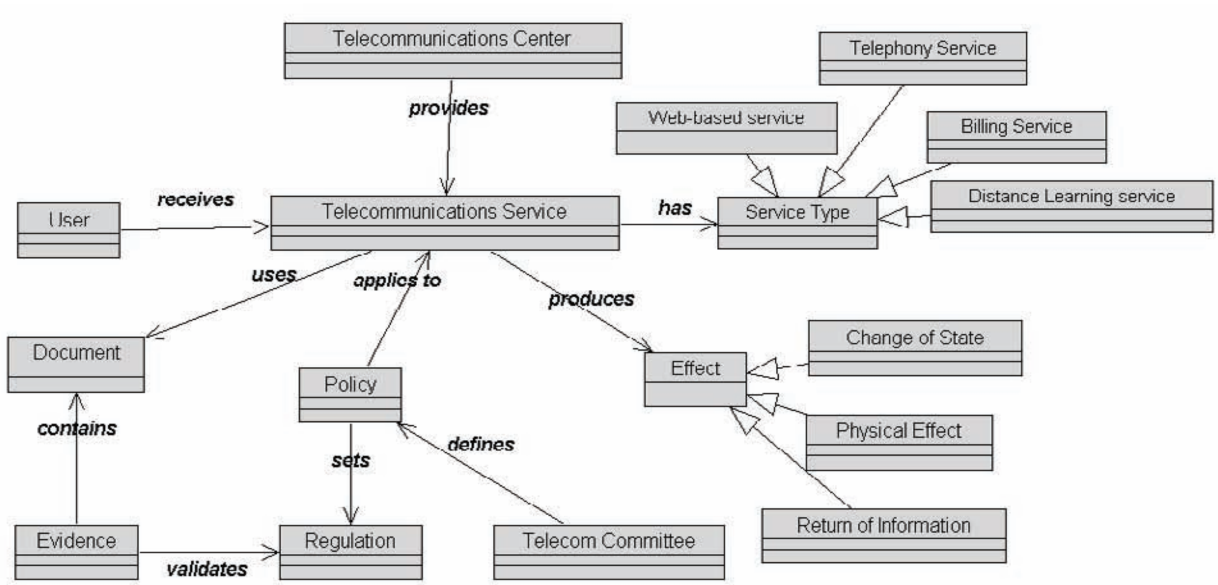
Data Model for Service Provision

Modeling of the service provision in the Public Administration domain is an attractive field for researchers (Goudos et al., 2007; Peristeras & Tarabanis, 2004). A telecommunications service object model is presented here. In order to explain the service model details, the administrative bodies that provide services and formulate the telecommunications policy in AUTh have to be given at first. The administrative body that formulates telecommunications policy in the university is the Networks and Informatics Committee. The Committee members come from both the academic and the administrative staff. The Committee chairman by default is one of the vice-rectors. The Committee proposes the telecommunications regulations (e.g., billing policy) for all university's staff and the new services to be deployed. The Telecommunications Center has been assigned the service provision and the service deployment. The Telecommunications Center staff is grouped into three categories: administrative, technical, and engineering.

From a service oriented perspective, the university's telecommunications service provision may be modeled using elements from Public Administration service object models like the Governance Enterprise Architecture (GEA) service object model (Goudos et al., 2007; Peristeras & Tarabanis, 2004) and the Open Group Service Oriented Architecture (SOA) (OpenGroup, 2008) reference model. The service provision object model is given as UML (Unified Modeling Language) class diagram as shown in Figure 3. A brief explanation of the objects shown in the diagram follows.

The actors involved in the service provision are the *Telecommunications Center,* which plays the role of the service provider, the *Telecomm Committee*, which defines university's telecommunications *Policy*, and the *User* that receives the service. The *User* in this case is the network subscriber (whole university's staff) that is entitled to telecommunications services. The *Policy* sets the *Regulation* for service execution. This *Regulation* is validated by *Evidence* contained

Figure 3. Telecommunications service basic data model



in administrative *Documents* that are used as the service input. For example, a regulation defines that a professor is entitled to a digital telephone set free of charge. When a professor applies for this service submits a document that contains his election act as a university professor. This is the *Evidence* that validates the *Regulation*. The service produces output that in accordance with the Open Group and the OASIS SOA model (OASIS, 2006) is called an *Effec*t. A real world *Effect* may consist of a *Change of State*, *Return of Information* and a *Physical Effect*. A *Change of State* declares that the service execution has resulted in a change in the information space. For example, the installation of a new telephone set with new phone number changes the information state of a phone database. The *Return of Information* is another possible effect (e.g., a DVD containing a distance learning session held). A *Physical Effect* is the actual change in the real world (e.g., the client receives a new telephone set). A *Telecommunications Service* has *Service Type*. The *Service Type* can be any of the four categories: *Telephony Service, Billing Service, Distance Learning Service* and *Web-based Service*.

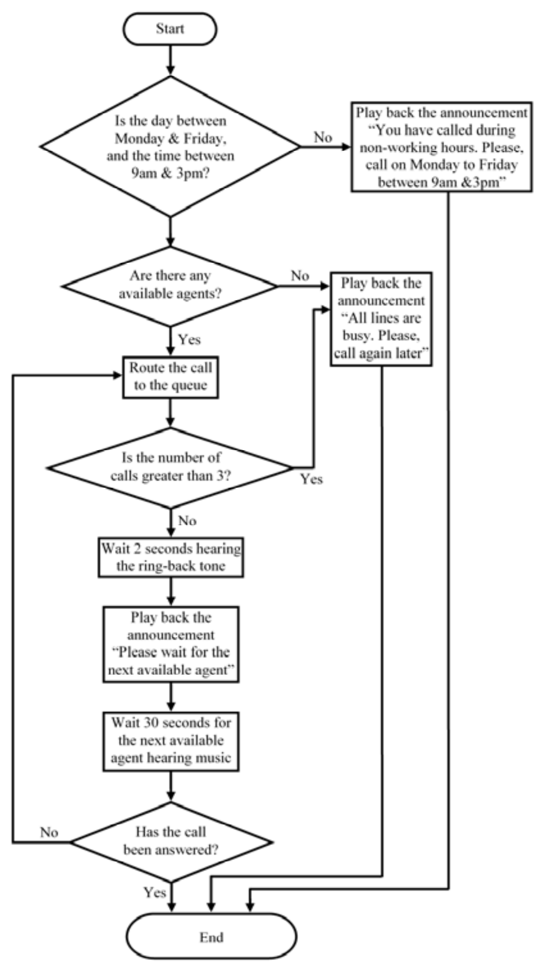
Basic Services Offered

An important issue for every large network is the operation of a daily help-desk service. The help-desk service plays the role of an interface between users and services. The help-desk service combines the advanced telephony features of a call center with a number of skilled agents. The help-desk call center operations flowchart is given in Figure 4. The agents answer to telecommunications related questions and provide assistance for service requests. A web site is also available (http://www.tcom.auth.gr) for finding information about services, searching an on-line telephone directory, finding service application forms and user guides for several telephone devices, etc.

As stated, there are four service types. Telephony services include call center activities, Interactive Voice Response (IVR) applications, ISDN Basic Rate connections, digital telephony features (e.g., group communication, malicious call detection, handling incoming calls, routing outgoing calls etc.), creating business groups (e.g., grouping extensions that belong to people working together so that any member of the group can answer to a call in any group extension) and voice mail. The Billing services include the issuance of bimonthly bills and videoconferencing bills. Distance learning services are services closely associated with the university's educational procedure. Table 2 summarizes.

The generic service provision UML statechart is given in Figure 5. The initial state is that of

Figure 4. Help-desk call center operations flowchart



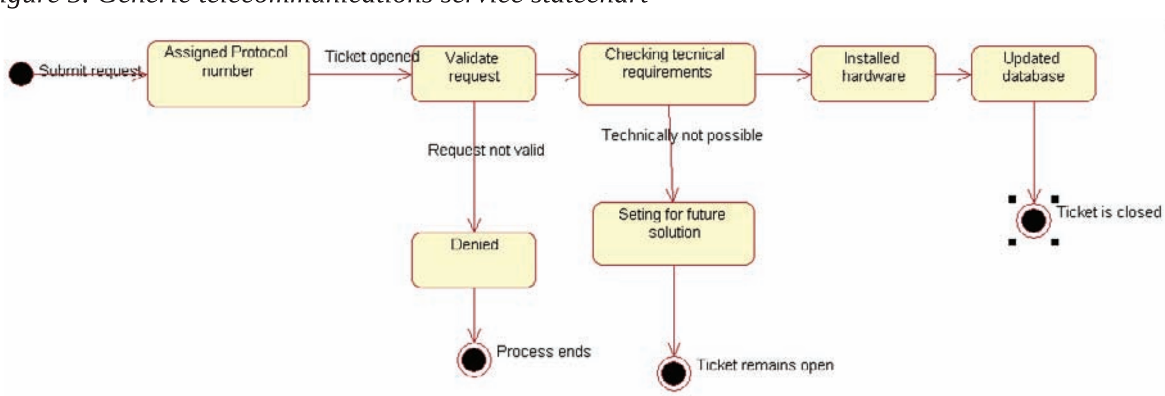
an application form for service request. The administrative staff assigns a protocol number and stores the request data. A ticket is opened meaning that the request is in process. The request is validated. If the user is not entitled to this service the process ends and a denial document is produced. If the request is valid the technical requirements to fulfill it are being checked. If the request cannot be implemented for technical reasons the ticket remains open waiting for a future upgrade. For example a new phone request cannot be completed if there is not a structured cabling path between the working place and the building's main distribution frame. Whenever it is necessary to satisfy the request, new hardware is installed. Information is updated in all related databases. Finally, the process is completed and the ticket is closed.

Table 3 has the statistics of service requests completed successfully in 2007. The digital telephony service requests represent the 54% of the total service requests. This can be explained easily due to large number of new connections installed every year in AUTh's network. The average service execution time was about 1 hour for web-based services, 3 hours for billing services, 1-2 days for digital telephony services (this varies depending on the service type) and 1 day for distance learning services. It must be pointed that out that there is an increasing demand for

Table 2. Basic services offered

Digital Telephony services	**Billing services**	**Distance learning services**	**Web-based services**
ISDN Basic Rate connections	Issue bimonthly telecommunications bills	Group videoconferencing	Search on line phone catalog
Interactive Voice Response (IVR) applications	Issue videoconferencing bills	Multipoint Videoconferencing	Search on-line bills database
Voice mail		Digital content creation and maintenance	Find application forms and documents
Personal Authorization Code			Video streaming
Malicious call detection			Discussion forum for distance learning and telephony services
Call center service			Distance learning classroom book keeping application
Advanced features (multiparty conference, business groups, etc.)			Online help-desk

Figure 5. Generic telecommunications service statechart



advanced services during the last years. Publicity actions like emails to all staff, the making of daily seminars for new services and printing leaflets have helped to increase this demand.

An important and demanding task is the distance learning service provision (Lange, 1994). As mentioned previoulsy, the Telecommunications Center provides distance learning services to all academic staff. The group videoconference systems installed in these classrooms provide full duplex real time audio and video connections with remote sites over ISDN (Jacobs & Rodgers, 1998) (H.320) or IP (H.323) networks. Apart from point-to-point connections, a Multipoint Control Unit (MCU) is also available, allowing up to eight simultaneous videoconference connections. Thus, communication and cooperation among educational institutes or organizations become easier. Telecommuting application or virtual meeting is also an option.

The distance learning service provision is further assisted with a set of supplementary services (Frieden, 1999). A distance learning web site has been created. This site is updated on a weekly basis. It contains information about the distance learning process and presents in detail all the equipment used. A useful feature is the distance learning map of Greece. Technical details and links to all distance learning facilities in Greek Universities can be found in that map. Several short videos about distance learning sessions held can be downloaded.

Every distance learning session held in the classrooms can be broadcasted live in the Internet. This is accomplished using video-streaming technology (Ahmad, Wei, Sun, & Zhang, 2005) through the aforementioned site. Additionally, a digital record of lectures has been created and a DVD copy can be given to anyone interested. The whole lecture videos in a high compression format (e.g., Microsoft wmv files) can be downloaded on demand.

Table 3. Services offered statistics for year 2007

Digital telephony services	Billing services	Distance learning services	Web-based services
54%	25%	9%	12%

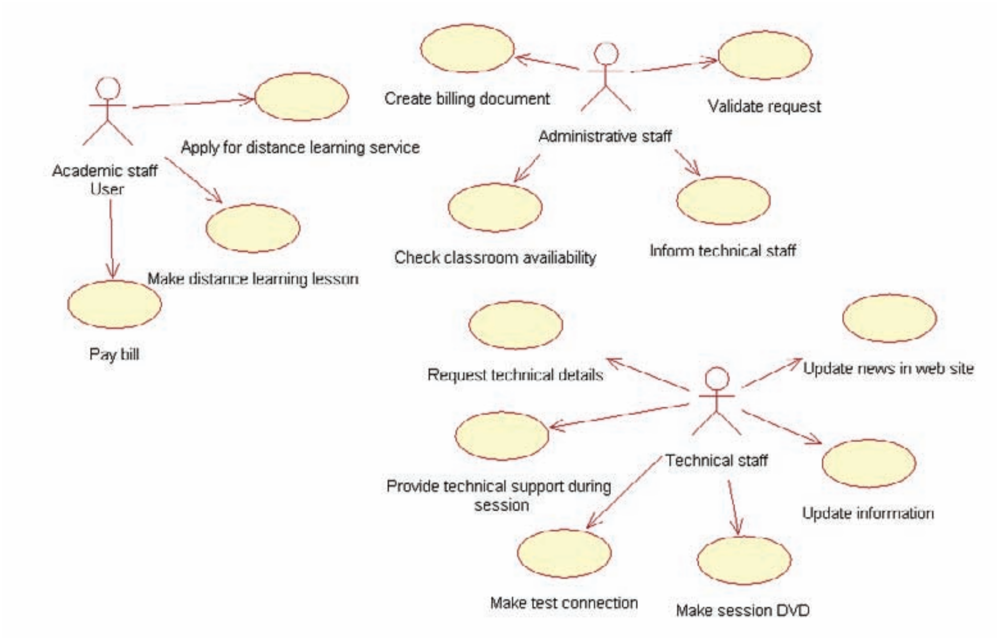
A voice portal has also been created using the advanced telephony services offered by the university's PBX. Anyone interested in distance learning may hear a number of recorded messages or talk directly to technical personnel. The portal offers information about the distance learning facilities in the classrooms and the digitalization of the lecture held afterwards.

The generic distance learning service use case is rather complicated. This is due to the academic nature of AUTh. We believe that this use case can be of particular interest therefore we describe the service details as follows. This use case is given in a UML use case diagram of Figure 6. It involves three actors; the user (academic staff member), the Telecommunications Center administrative staff and the Telecommunications Center technical staff. The user applies for service provision using a predefined application form. The administrative has to validate his request, check classroom availability, inform the technical about distance learning session details and finally issue a billing document after the session is completed. The technical staff contacts a person in the remote site, requests the technical details (type of equipment, connection speed, etc), and makes a test connection to ensure proper equipment functioning.

Billing Policy

One of the most difficult, yet important, issues is to develop a billing policy. The university's Senate is the administrative body that votes the billing policy proposed by the Networks and Informatics Committee. The university's policy has two goals. The first is to offer advanced telecommunications services in order to further assist research and educational duties of its staff. The second is the provision of some of these services free of charge.

Figure 6. Distance learning use case diagram



Most of the advanced telephony services described previously are offered free of charge.

For example, calls to local destinations are free of charge for all university's staff but long-distance, international and mobile calls can be made only by using a valid PAC. This PAC is charged with the call bill. Calls using a PAC can be made from any telephone set in any network node, thus allowing per person charges instead of per extension. Every university staff member is entitled to an amount of free units in a two-month period. When the PAC charging exceeds this free amount charges appear on a bill for this staff member. The number of free units depends on the user's position in the university's hierarchy. For example, professors are entitled to a larger amount of free units per month than lectures or administrative staff. The billing period is set to two months.

The distance learning services follow another billing policy. Technical support is provided free of charge to all university's staff for sessions held during working hours. The user has to pay only the call charges in case of ISDN (H.320) videoconferences. In case of videoconferences over the Internet (H.323) no call charges exist. A different billing policy exists for external users that do not belong to the university's staff. In this case, there are call and technical support charges.

NETWORK PLANNING FUTURE TRENDS

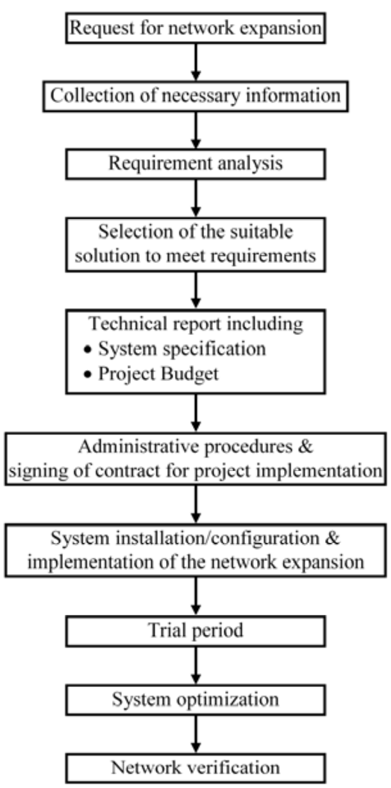
In this section we present four critical parameters of the network planning process aiming to provide valuable guidelines to network administrators.

Expansion Procedure

The university is a live organization that expands constantly with the foundation of new undergraduate or postgraduate departments and the relocation of existing departments to new buildings. The network expansion is a complex task, mainly because it depends on several factors concerning the new location. Therefore, a well-designed and fully documented expansion procedure is necessary in order to assure that the network will meet successfully the user requirements imposed. The expansion procedure covers several issues as shown in Figure 7.

A request for voice and data network in a new location is submitted at first from the corresponding university department. Therefore, a large amount of information must be collected concerning several parameters, such as the geographic location of the new building and its distance from the existing network nodes, the way the building space is distributed into working areas per floor, the number of working seats per area and finally any requests for specific telecommunications equipment (e.g., an alarm system, surveillance cameras, videoconference systems, specific laboratory equipment like remote sensors). Taking into account the aforementioned factors, a requirement analysis takes place. This analysis leads to the selection of the most suitable solution that meets the given requirements.

Figure 7. Expansion procedure diagram



An important factor in large networks is the wiring installation (Budenski, 2004; Semenov et al., 2002). The structured cabling installation requirements have to be documented in the new building. Therefore, a technical report is created first about the wiring requirements. A technical report about the telecommunications active components that have to be installed follows. The specifications of the selected system as well as the procedure of implementing the network expansion are described and analyzed in these technical reports. The total budget of the proposed solution has to estimated and described in detail in the aforementioned reports.

Afterwards, the appropriate administrative procedures are applied in order to find one or more contractors who conform to the technical reports. After the signing of the contract for the project implementation, the contractor installs and configures the system and completes the whole network expansion. In case of structured cabling installation, the contractor has to fully document and measure the new location.

Then, a trial period starts in order to find out problems caused by parameters and requirements not taken into account during the collection of necessary information or during the requirement analysis. An effort to solve these problems is made aiming to optimize the system. The final step is to verify the network functionality by monitoring traffic reports for congestion problems.

Cost Reduction Case Studies

The telecommunications costs are a significant part of the university operational costs. Network planning has to take into account not only the installation costs but also the operational costs. Therefore cost reduction is always a priority for network management.

Two major factors affect the university's cost reduction policy. The first is the deregulation of the telecommunications market in the European Union (EU) and Greece in 2001 (Psimmenos), which has resulted in an increasing number of telephone companies competing for cheaper services. The second factor is the increasing number of mobile phone users in Greece during the last decade. The mobile calls percentage among university users has grown from 5% to about 20% of the total calls. Figures 8a and 8b show the call statistics for the year 2007.

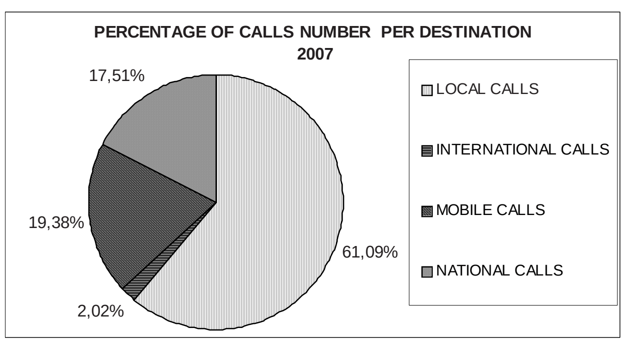
The Telecommunication Center staff has recently prepared a technical report for reducing AUTh's telecommunications costs. Among the report's conclusions was the fact that the utilization of emerging technologies for cost reduction is not always a trade-off with reduced service quality. The report proposed the usage of existing technologies and new market offers for cost reduction.

In Greece, a leased analogue line (M.1020) costs about the one tenth of a digital E1 line (2.048 Mbps). Therefore, the proposal for cost reduction requires the replacement of costly digital leased lines with cheaper analogue lines combined with HDSL or SDSL modems that connect at E1 speeds. The rise of VoIP technology has also helped in reducing costs. The connection with other Greek universities through a VoIP gateway provides toll free calls to these destinations. The routing of outgoing calls over the Internet is also planned.

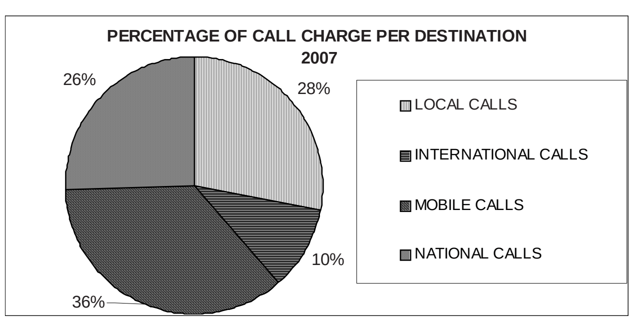
New offers and service providers have appeared in the Greek telecommunications market over the last years. For example, mobile service providers have introduced lower rates for enterprises and organizations. Thus, a significant action towards cost reduction was to make contracts with mobile service providers. Fixed Cellular Terminals (FCTs) have been installed in the university's premises and all outgoing calls to mobile users are directed over these lines. This particular action has caused a 70% drop in mobile phone charges.

Greek telephone companies have also offered reduced cost packages for international destina-

Figure 8. Call statistics for the year 2007



(a)



(b)

tions. Such a package was chosen for university's ten most popular international call destinations. In this way, the university has saved about 30% of the amount paid for international calls. Recently, a new contract was signed between the university and a major Greek telecommunications company regarding local and long distance calls. The predicted savings from this action may come up to 30% of the total amount of bills paid.

From the previously stated, one may conclude that cost reduction in a telecommunications network is a continuous effort that depends on emerging technologies and market conditions.

AUTh's Telecommunications Network - SWOT analysis

SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis could be a valuable strategic planning tool. A SWOT analysis concerning the AUTh Telecommunications Center is presented. Strength is a resource that the Telecommunications Center can use in order to achieve its objectives. Weakness is any internal limitation that could prevent Telecommunications Center from achieving its goals. Opportunity is any favourable situation in its environment. Threat is any unfavourable situation in the academic environment that could potentially cause problems to the implementation of the Telecommunications Center's strategy. Table 4 presents the described factors.

From the characteristics given in Table 4, it appears that the Telecommunications Center's strengths concern both its human resource and the network infrastructure. The skills and the capabilities of the workforce are key factors that allow exploitation of the existing opportunities. The know-how of the workforce allows the Telecommunications Center to create a development plan that will convert the existing network to a fully-IP enabled network. Over the last decade, the EU has a financing policy for ICT research projects concerning both the development of new technologies and the provision of new services. This policy may help to implement the aforementioned plan. The staff shortage weakness has been caused due to the lack of flexible procedures for new staff employment and the lack of adequate funds. Thus, the EU financing could help overcoming this weakness. Due to the complex nature of the network infrastructure and technologies, the adaptation and training of the new staff requires a long time period.

This optimistic view of the network evolution could be obstructed by the subscriber behaviour. Statistical surveys have revealed that the fear against new technologies is a crucial factor that could inhibit any upgrade procedures. In order to overcome this obstacle, the Telecommunications Center exploits the worldwide presence of VoIP and the deregulation of the Greek telecommunications market that make the new technologies

Table 4. SWOT analysis for the telecommunications center

Strengths	• Highly skilled workforce • Lower response time with efficient and effective service • Extended Network Infrastructure covering all University's facilities • Several years of experience
Weaknesses	• Network Infrastructure not fully IP-enabled • Staff shortage • Adaptation and training of the new staff requires a long time period
Opportunities	• Technological evolution and worldwide presence of VoIP • European Union's policy for financing Information and Communications Technology (ICT) research projects • Deregulation of Greek telecommunications market
Threats	• Subscriber's fear against the new technologies • Greek Public Sector's legislation and bureaucracy • Lack of flexible procedures for new staff employment

not only available but also familiar to anyone. Moreover, this threat reveals the need for change management. Educational programs can help subscribers adopt the new technologies.

From the aforementioned description it is clear that the Telecommunications Center has the ability to exploit the existing opportunities in order to overcome its weaknesses and face its threats. The only threat with unpredictable results is the Greek Public Sector's legislation and bureaucracy. This threat requires a good preparation not only of a development plan but also of an alternative plan that will be used according to the case.

The aforementioned analysis has helped the Telecommunications Center to evaluate its future directions and has proved to be a helpful tool in the migration analysis, which is discussed as follows.

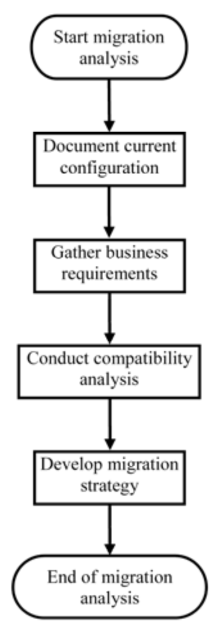
Migration Analysis

The technological evolution and the VoIP presence drive AUTh to the next network upgrade step. The migration to a fully IP-enabled converged voice and data network is the obvious choice. The benefits of adopting a fully IP-enabled communications network over the traditional TDM-based approach have been given in (AVAYA, 2005; Buckley, 2005). The main objective is to ensure an efficient and trouble-free migration of the telecommunication network.

To facilitate the migration task a migration analysis must be carried out. This analysis is composed of the steps presented in the flowchart of Figure 9. The first step is to document all the current equipment details like circuit pack numbers and versions, software and hardware versions of nodes and the current user equipment configuration. The documentation of in-use telephony features in every node is also important. The next step is to gather the business requirements. This step includes the definition of constrains and requirements. The network software and hardware dependencies have to be determined. Priorities have to be assigned for all running applications and possible vendor problems have to be identified. The compatibility analysis follows, where the current network and the user equipment are analyzed in terms of hardware and software compatibility. Compatibility issues in service provision are also analyzed. The final step is to develop a migration strategy. This step includes teamwork to make an early migration strategy draft, review and update it. The downtime due to migration has to be evaluated. All affected university's departments and schools have to be notified. Finally, a revised version of the original strategy is prepared.

Several technical issues will have to be resolved in a fully IP architecture. These can be grouped into management and security related issues. The elaboration of a new security policy is necessary. Security issues like threats from DoS attacks or worms have to be considered. The usage of private IP addresses in both signalling and voice connections combined with Network Address Translation (NAT) may provide a possible solution. Enabling firewalls and IPsec in central nodes is another option. The combination of both

Figure 9. Migration analysis flowchart



previously stated practices will be the main part of a future security policy. The management issues include usage of new management protocols. The Network Configuration Protocol (NETCONF) defined in RFC 4741 provides mechanisms to install, manipulate, and delete the configuration of network devices. NETCONF uses XML for data encoding purposes and it is independent of the data definition language and data models used to describe configuration and state data. The COPS (Common Open Policy Service) Protocol (RFC 2748, updated in RFC 4621) describes a simple client/server model for supporting policy control over QoS signalling protocols. The exploitation of the usage of protocols like NETCONF and COPS will be part of our future work. Multihoming optimization is another important and complex issue that will arise in new fully IP enabled network. Network preparation should also address this issue.

Despite the technological advantages, the migration will lead to new business models about the service provision. New value-added services will be deployed that will allow further call cost reduction for the academic staff. For example, applications like soft phones running in PCs will offer the features of an expensive digital set thus providing advanced telephony services for all. The usage of VoIP in campus Wi-Fi networks will offer a cheaper alternative to cell phones in campus. The integration of data and voice will give a boost to web-based services combined with IVR technologies like VoiceXML (Anderson et al., 2001). The adoption of new technologies forces changes in business processes, therefore making organizational changes management increasingly important. The most important factor is to communicate throughout the organization the way that the technological changes effects will support organizational goals.

CONCLUSION

We have presented the case study of the telecommunications network of the Aristotle University of Thessaloniki, Greece. The academic environment and the size of the organization have imposed particular demanding and complex requirements in network management and planning. Examples of periodic network management procedures have been given. The high network availability and toll fraud prevention are the most important management tasks. A data model for telecommunications service has been presented and explained using a SOA perspective. The statistics show that the average service completion time ranges from an hour to few days depending on the service type. We have found out that it is very important to make the users aware of the services and service capabilities. This goal can be achieved through a series of informative actions. The billing policy is also important because it helps funding services and equipment. Network planning is a continuous effort that takes into account both the economic and the technological environment circumstances and conditions. Network expansion has to be a fully documented and well-defined procedure. It has been found that the utilization of emerging technologies for cost reduction is not always a trade-off with reduced service quality. Technology combined with telecommunications market offers can lead to significant cost reduction. SWOT analysis has the potential to reveal the key elements that have to be taken into account for network planning. A migration analysis is the key factor for the adoption of a successful strategy. Technological benefits have to be examined carefully and clearly explained to all involved parties. The cost benefits have to be estimated. Organizational and possible integration issues that may arise from the implementation of the migration strategy have to be resolved. The aforementioned show that network planning issues can be viewed from several different perspectives. The correct perspective depends strongly on the organization type. In an academic environment the user's needs for telecommunications services are in general different than those in other kinds of organizations. Academic related factors like research and education influence directly the structure, the management and the planning of a university's telecommunications network.

REFERENCES

Ahmad, I., Wei, X., Sun, Y., & Zhang, Y. Q. (2005). Video transcoding: An overview of various techniques and research issues. *IEEE Transactions on Multimedia, 7*(5), 793-804.

Anderson, E., Breitenbach, S., Burd, T., Chidambaram, N., Houle, P., Newsome, D., et al. (2001). *VoiceXML Early Adopter* (1st ed.): Peer Information Inc.

Atkins, J. (2004). New tools for new networks. *Telecommunications International, 38*(8), 16-20.

AVAYA. (2005). Migrating to Converged Networks and IP Telephony Applications:Implications for Management Tools and Managed Services *White Paper.*

Azevedo, H. L. C., Vellasco, M. M. B. R., & Passos, E. P. L. (2001). Data mining applied to CRM for a long distance telephone company. *Advances in Neural Networks and Applications*, 327-332.

Bollapragada, R., Morawski, T. B., Pinzon, L. E., Richman, S. H., & Sackett, R. (2007). Network planning of broadband wireless networks. *Interfaces, 37*(2), 143-162.

Buckley, S. (2005). Governments grasp for IP communications. *Telecommunications (Americas Edition), 39*(11), 23-24.

Budenski, D. (2004). Keeping up with the evolution of telecommunications wiring. *EC and M: Electrical Construction and Maintenance, 103*(8), 20-24.

Douskalis, B. (2000). *IP Telephony: The Integration of Robust VoIP Services*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Engebretson, J. (2007). Voice's place in the fiber future. *Telephony, 248*(9), 26-31.

Frieden, S. (1999). Support services for distance education. *Journal of Educational Technology and Society, 2*(3), 77-88.

Goudos, S. K., Loutas, N., Peristeras, V., & Tarabanis, K. (2007). *Public Administration Domain Ontology for a Semantic Web Services EGovernment Framework*. Paper presented at the IEEE International Conference on Services Computing, (SCC 2007)

Hilas, C. S., Goudos, S. K., & Sahalos, J. N. (2006). Seasonal decomposition and forecasting of telecommunication data: A comparative case study. *Technological Forecasting and Social Change, 73*(5), 495-509.

Jacobs, G., & Rodgers, C. (1998). ISDN-based distance learning: Tangible costs, intangible benefits. *Computers and Education, 31*(1), 41-53.

Kuhn, D. R. (1997). Sources of failure in the public switched telephone network. *Computer, 30*(4), 31-36.

Lange, J. (1994). ISDN videoconferencing for education and training. In R. Mason, Bacsich, P. (Ed.), *ISDN Applications in Education and Training* (pp. 127-143). London: IEEE Press.

Lee, I. (2005a). An analytical model of e-recruiting investment decision: An economic employment approach. *IEEE Transactions on Engineering Management, 52*(4), 486-496.

Lee, I. (2005b). The evolution of e-recruiting: A content analysis of fortune 100 career web sites.

Journal of Electronic Commerce in Organizations, 3(3), 57-68.

McCormick, A. (2007). Service-enabled networks aid convergence. *Lightwave, 24*(5).

Medhi, D. (2007). Routing management in the PSTN and the Internet: A historical perspective. *Journal of Network and Systems Management, 15*(4), 503-523.

Mezher, T., Abdul-Malak, M. A., & Dayya, M. (1999). Decision support system for managing infrastructure telecommunications projects. *Journal of Management in Engineering, 15*(1), 74-81.

OASIS. (2006). Reference Model for Service Oriented Architecture 1.0. *Committee Specification 1*, Available from http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf

Okuno, S., & Okuda, M. (2003). Trends toward Globalization of the TL 9000 Quality Management System for the Telecommunications Industry. *NTT Technical Review, 1*(6), 92-95.

OpenGroup. (2008). Service Oriented Architecture (SOA). Available from http://www.opengroup.org/projects/soa/

Peristeras, V., & Tarabanis, K. (2004, 25-27 October). *Governance Enterprise Architecture (GEA): Domain Models for e-Governance.* Paper presented at the 6th ACM International Conference on Electronic Commerce (ICEC'04), Delft, The Netherlands.

Prabhakar, G., Rastogi, R., & Thottan, M. (2005). OSS architecture and requirements for VoIP networks. *Bell System Technical Journal, 10*(1), 31-45.

Pravda, I., & Vodrazka, J. (2007). Voice quality planning for NGN including mobile networks. *IFIP International Federation for Information Processing, 245*, 376-383.

Psimmenos, I. (2007). The EC single market and employees' participation. Experience from the telecommunication and energy utilities. *Utilities Policy*, 24-37.

Schulzrinne, H., & Rosenberg, J. (1999). Internet telephony: Architecture and protocols - an IETF perspective. *Computer Networks, 31*(3), 237-255.

Semenov , A. B., Strizhakov, S. K., & Suncheley, I. R. (2002). *Structured Cable Systems* (1st ed.): Springer.

Uppaluru, P. (2003). Discover the power of voice: Transforming the telephone to open new business models. *Alcatel Telecommunications Review*(3), 329-334.

Van den Ende, J. (2003). Modes of governance of new service development for mobile networks: A life cycle perspective. *Research Policy, 32*(8), 1501-1518.

KEY TERMS

Busy Hour Call Completion (BHCC): The number of calls that a telephone system can complete during the busy hour of the day. BHCC is a measure of system processor capacity and a factor considered in traffic engineering.

Call Data Record (CDR): Records generated by subscriber traffic and used later to bill subscriber for service.

Fixed Cellular Terminal (FCT): (also called Premicell or GSM Gateway) Device enabling a GSM SIM (Subcriber Identity Module) to be utilized from a fixed line handset as though it was calling from a GSM mobile phone. Using a fixed handset with a FCT, users make and receive calls through a GSM network. Usually, such calls are cheaper than those using a normal landline carrier.

H.320: Umbrella recommendation from the ITU-T for transmitting audio, video and data signals over ISDN-based networks.

H.323: Umbrella recommendation from the ITU-T for transmitting audio, video and data signals over packet-based networks.

High Bit-Rate Digital Subscriber Line (HDSL): One of the earliest forms of DSL used for high bit-rate (2Mbps) between two endpoints over two pairs of copper wire. HDSL is a symmetrical line meaning that equal amounts of bandwidth are available in both directions.

Interactive Voice Response (IVR): Technology that allows a computer to detect voice or touch tones coming from a normal phone call.

Symmetric Digital Subscriber Line (SDSL): A Digital Subscriber Line (DSL) variant with 2 Mbps data rates. It runs over one pair of copper wires.

Time Division Multiplexing (TDM): A multiplexing type where signals from different sources are transmitted over the same physical link but in different time intervals called timeslots.

VoiceXML: The W3C's standard XML format for specifying interactive voice dialogues between a human and a computer. It allows voice applications to be developed and deployed in an analogous way to HTML for visual applications.

Chapter XLI
Planning of Wireless Community Networks

Károly Farkas
University of West Hungary, Hungary

Csaba A. Szabó
Budapest University of Technology & Economics, Hungary

Zoltán Horváth
Budapest University of Technology & Economics, Hungary

ABSTRACT

The objective of this chapter is to discuss the applications, state-of-the-art technologies, planning methods and business models for wireless community networks and provide an integrated presentation of these essential parts with examples. After a short introduction we give an overview about the state-of-the-affairs of wireless community networks presenting the driving forces, stakeholders, services and built upon applications in some carefully selected projects. Then, we discuss the technological aspects and suggest a design methodology for planning wireless community networks. The application of this methodology is illustrated via an ongoing digital city project in Hungary. After this we analyze the relevant business models. And finally, before concluding the chapter we give a short outlook discussing the future trends of the area.

INTRODUCTION

Broadband access to citizens, communities, public institutions and developing businesses is a natural need today. However, there is no a straightforward way to provide broadband access to these users. Although today's technologies (from ADSL via fiber-to-the premises to broadband wireless access) are appropriate for this purpose, telecommunication companies cannot afford their deployment and serve these users due to the large costs and very long return on investment (ROI) period. As an alternative, a so-called community network (CN) can be created and deployed by governments and international organizations worldwide serving strategic purposes, too. Community networks are able to foster economic development, help start-ups grow, bring new businesses into the region and can alleviate also the problems related to the digital divide. Reflecting these objectives a huge number of community network related initiatives have been launched recently in North America as well as in Europe.

Community network infrastructures can be built by using the common technologies in telecommunication companies' networks. Fiber has been an attractive solution for many cities, first of all in North America. Building a fiber network is technically viable where a local government or some of its utility companies own ducts and support structures forming "free" assets. However, for economical feasibility it is necessary to have a few large customers, such as ISPs (Internet Service Providers), which use and pay for a substantial share of fiber capacity. On the other hand, wireless technologies (e.g., Wi-Fi, WiMAX or 3G mobile) are almost always appropriate for building community networks because they are relatively easy to install, expand and operate in a cost effective manner.

Planning, deployment and operation of community networks have been challenging tasks. As opposed to telco networks, there is a specific set of services that the city or region wants to implement. These services and built upon applications are to be made accessible for a wide range of geographically diverse users regardless their locations. Moreover, cities can more freely choose communication technologies, including emerging ones, as they do not have the stringent business requirements the telecommunication companies have to meet, such as short ROI or totally risk-free adaptation of new technologies. However, suitable business models have to be defined with clever constructions of involving both the public and private sectors, while satisfying legal and regulatory requirements.

The objective of this chapter is to discuss the applications, state-of-the-art technologies, planning methods and business models for wireless community networks and provide an integrated presentation of these essential parts with examples. This integrated approach has been rarely found in the technical literature. Our previous works (Szabó et al., 2007; Szabó et al., 2008) address the issue of wireless community networks. A recent edited book (Chlamtac et al., 2005) attempts to bring together the most important aspects – technical, legal, regulatory and economic – of community networks into one book. Within the framework of an ongoing European Network of Excellence project called OPAALS (OPAALS, 2008), the social side, information technologies and economic models are being investigated by a large inter-disciplinary international team, also putting community networks in a wider context of digital ecosystems and digital business ecosystems.

As for the related work, not only the aforementioned integrated approach is missing in the existing literature, but also different aspects of the theory and practice of community networks are dealt with in unequal depths. Social aspects are well treated, there is a significant amount of sociology literature focusing on socio-technical infrastructures, see e.g. Star and Ruhleder (1996), Edwards et al. (2007). As far as the applications are concerned, the paper by Intel is worth mentioning first of all (Intel, 2005) which points out to some common core technologies that are critical for the development of digital community services. Other papers deal with the design of particular society-related services, such as e-learning or telemedicine. There is a significant amount of literature on business models, including general considerations, regulatory and competition aspects on one hand, and a few high level, not detailed analyses of specific applications such as automated meter reading (AMR), see Settles (2007). A complete business model and calculations of a wireless community network cannot be found. As for the technology planning, the application-driven methodology, introduced in this paper, is not treated in the existing literature. Wireless network planning can be broadly divided into topology planning and QoS planning. As for the latter, the paper Niyato et al. (2007) is worth mentioning which addresses the issue of MAC layer design of WiMAX taking into account of requirements of telemedicine applications. However, to the best of the authors' knowledge, there are no publicly available tools for Wi-Fi mesh and WiMAX topology planning.

BACKGROUND

In this section, first we discuss the stakeholders, initiators and the set of applications usually considered for community networks, then present one European and two North American case studies as representative examples of community network projects.

Community Networks

Under the term "community network" or CN we mean the combination of the telecommunication infrastructure created by the participation of the local government or public organization, the services provided upon it and the specific business model to operate the infrastructure and provide services. Here we mainly focus on wireless CNs, where at least the access part but in many cases also the distribution and backbone parts are implemented using a wireless and/or mobile technology.

The stakeholders of community networks include: (i) public agencies (local governments, local development agencies, public organizations); (ii) users (citizens, SMEs (Small and Medium Enterprises), associations, etc.); (iii) private sector service providers (e.g. telcos or ISPs); (iv) local and global facilitating agencies (such as research and consulting centers, associations of community networks). Usually one of the aforementioned stakeholders is the "initiator" of the project. Classical community networks were initiated by the communities themselves as grass root initiatives, while most of today's projects are planned and implemented by some form of local and/or regional governments' participation.

Applications that drive the development of community networks can be grouped as follows:

A. Access to public information and services
   - Public Internet kiosks for access to public information, e-government services, tourism.
   - Portals for e-government services, for local communities and for tourists.

B. Public safety
   - Enhancing public safety by remote surveillance of public areas.
   - Improving the communication with police, civilian police, fire department and alike.

C. Traffic control and transportation
   - Coping with traffic congestion by vehicle monitoring and intelligent traffic light control.
   - Vehicle management for public transportation (buses).
   - Intelligent parking systems with flexible payment.
   - Monitoring of road conditions, in particular in winter.

D. Health care
   - Improving the efficiency and cost-effectiveness of health care services by broadband and wireless communications among and within health care providers.
   - Providing telemedicine services.
   - Home health care and assisted living.

E. Business services
   - Business partners/providers/clients searching.
   - Digital services: search, use and combine.
   - B2B (Business to Business) and B2C (Business to Client) transactions.
   - Advertise product and services.

F. Educational
   - Remote classroom.
   - Remote consultation.

G. Utility companies
   - AMR (Automated Meter Reading) for electricity, water, gas companies.
   - On-line maps with pipes and wires helping outdoor works.

In most cases, there is usually 1 or 2 applications that are the main motivations for the implementation of a given community network. Below we list some wireless CN initiatives together with their primary applications (Intel Solutions, 2005):

- **Chaska, MN, US:** Digital divide for schools, businesses and residents;
- **Cheyenne, WY, US:** Traffic signal management;
- **Corpus Christi, TX, US:** Automated meter reading for city-owned utilities;
- **Lewis & Clark County, MT, US:** T1 replacement; access to remote county buildings;
- **Ocean City, MD, US:** Integrated digital, voice and video for city buildings;
- **Piraí, Brazil:** Municipal field-force productivity; promotion;
- **Portsmouth, UK:** Bus passenger information dissemination;
- **San Mateo, CA, US:** Police field-force productivity improvement;
- **Shanghai, China:** Police field-force productivity improvement;
- **Spokane, WA, US:** Municipal applications and e-Government initiatives;
- **Westminster, UK:** Video surveillance and enhanced security.

Case Studies of Recent Community Network Projects

As of August 1, 2007 there were 92 regional and city-wide networks, 68 city hot zones and 40 public safety and municipal use networks, alone in the US (source: MuniWireless (2008), one of the well-known portals of wireless community networks). Furthermore, 215 city and county-wide projects are in progress. The total number of existing networks and ongoing projects is 415, which reflects an exponential growth regarding the figure (122) two years ago. There are similar initiatives around the globe and a comparable growth is expected to happen in the next few years. Although Europe, at least the continental part, seems to be lagging behind the US, the ambitious European plans to penetrate broadband services to citizens and institutions and foster regional development can help the spreading of CNs even on the old continent.

Below we present three case studies of recent community network projects that represent different objectives, target applications, stakeholders and business models. Although all these projects aim at providing various services and applications, in each case there is one primary application on which the business model is built. *T.Net* in Italy aims at creating a telecommunication infrastructure in a province that is sparsely populated and geographically challenged. The main goal of *Wireless Philadelphia* in the US is to provide Internet access in a city where the Internet penetration is quite low, while the primary objective of the *Corpus Christi* project in the US is to implement city-wide remote data collection for utility companies.

T-Net, Trentino, Italy

T.Net (Longano, 2007) is a community network project under implementation in Trentino, a province in Northern Italy. It is part of the eSociety initiative of the local government, whose strategic aims are: (i) the innovation of the local economy; (ii) the improvement of Public Administration efficacy; and (iii) the reduction of the gap which keeps many citizens from participating in the Information and Knowledge Society. Its management model involves publicly controlled companies for the implementation and management of the broadband infrastructure, supplying of transport services, connectivity and IT services for public administration and renting infrastructure to market operators under fair and non-discriminatory conditions. The network consists of a fiber optic backbone and a pre-WiMAX based (HiperLAN-2) wireless access network. The number of backbone nodes is 78 with the total length of optical cable over 750 km. The network connects 223 municipalities in total. Until the fiber infrastructure will be built, the province is leasing Gbit Ethernet facilities from Telecom Italia, the Italian incumbent telecom service provider. At the end of 2007, wireless access was provided for 150 municipalities.

Wireless Philadelphia, US

The Wireless Philadelphia initiative (The Wireless Philadelphia Executive Committee, 2005) started with a pilot, covering the central districts and is currently being expanded to cover the entire metropolitan area with a total of 20 million USD investments. The project is financed and implemented by EarthLink, one of the dominant ISPs in the US. The business model is built on providing Internet access in the city, as the level of broadband penetration is very low (below 25%) and is mainly based on dial-up access. EarthLink is also planning to sell bandwidth both to retail and wholesale customers. The city is planning to subsidize Internet access for low-income residents. Mobile workers that constitute half of the city workforce will communicate using this network infrastructure, supported by an already implemented Geo-spatial Information System (GIS). Other applications include video surveillance to reduce crime in the city.

Corpus Christi, US

In 2001, the city of Corpus Christi, which has about 250,000 inhabitants and an area of about 150 sq. miles, decided to implement an AMR system for water and gas customers (Corpus Christi Pioneers, 2007). The underlying network is an optical fiber backbone together with a Wi-Fi mesh network built and operated by Tropos Networks, the market leader in delivering metro-scale Wi-Fi mesh network systems. Overall, the city spent 20 million USD on the AMR system and on the wireless network, which yields a saving of 30 million USD over the estimated 50 million USD costs within the next 20 years without AMR. In addition to savings, the project resulted in higher level of customer service and support to citizens. After the rollout of the project, it was realized that the AMR application uses only a fraction of the bandwidth of the wireless network. Therefore, the city is planning to implement other applications including the support for public safety, health inspection, animal control, public works and utilities personnel.

TECHNOLOGICAL ASPECTS

In this section, first we give a short overview about the available wireless technologies on which wireless community networks can be built, then discuss the planning issues for technology selection regarding these networks.

Available Wireless Technologies

Below we overview three wireless technological alternatives such as Wi-Fi mesh, WiMAX and 3G cellular mobile. For further information interested readers should refer to any of the several books, e.g. (Webb, 2007), on wireless technologies.

Wi-Fi Mesh

Wi-Fi (Wireless-Fidelity) mesh networks are peer-to-peer multi-hop networks based on the IEEE 802.11 standard family (IEEE 802.11, 2008), where the nodes cooperate with each other to route information packets through the network. They present an alternative solution to "infrastructure based" networks like ADSL (Asymmetric Digital Subscriber Line). Mesh networks have some attractive features. Thus, they are "organic"; nodes may be added and deleted freely; the mesh principle means also fault tolerance, hence nodes may fail and packets will still be routed; mesh networks are manageable in a distributed manner. However, mesh networks also pose challenges. If there are too many nodes, the need for routing other nodes' traffic decreases the access throughput of a given node. On the other hand, if there are too few nodes then routing could be a problem. Security is also an issue. A practical problem is that today there are no interoperable products as the WLAN (Wireless Local Area Network) mesh standard (IEEE 802.11s) is relatively new. In spite of the aforementioned shortcomings, the majority of wireless CNs is Wi-Fi mesh and it is the most likely option to consider when someone is planning to create such an infrastructure. Current products feature dual/multiple radios (separate radio(s) for the access and backbone parts) to significantly compensate the throughput decrease when traffic is routed through a chain of nodes. Most recently combined devices have been also developed that implement both the Wi-Fi mesh and WiMAX capabilities using the latter technology for backbone purposes.

WiMAX

WiMAX (Worldwide Interoperability for Microwave Access) is an emerging wireless technology and a flexible telecommunication architecture based on the family of IEEE 802.16 standards (IEEE 802.16, 2008). It is often considered as the next generation of Wi-Fi networks, though the two technologies represent two different design and development lines from technical perspectives. The WiMAX technology provides a big step ahead evolution as the offered capacity and the communication range are approximately an order of magnitude higher compared to Wi-Fi. The topology of a WiMAX network can be point-to-point, point-to-multipoint or mesh. The area coverage is up to tens of km in LOS (Line Of Sight) environment at limited data rates. An attractive feature is operation under NLOS (Non Line Of Sight) conditions. Spanning only short distances, high capacity and data rates up to 100 Mbps can be achieved which make WiMAX a viable option for backbone and distribution network segments. It provides a high level of security due to the 3DES (Data Encryption Standard) and the AES (Advanced Encryption Standard) encryption standards. Quality of service is an inherent feature of WiMAX. It has several service classes including support for real-time data streams.

A typical fix WiMAX network architecture consisting of base stations, subscriber stations and different communication link types. To support mobile or nomadic users – implementing seamless handover of the user between the base stations – the mobile version of WiMAX based on the IEEE 802.16e standard was approved at the end of 2005, and products built on this standard have already been available. To deploy a WiMAX network is easy, quick and relatively inexpensive. Different spectrum allocation possibilities exist in licensed and license-free frequency bands. However, implementers of wireless CN infrastructures are cautious regarding WiMAX, mainly due to the currently high costs of WiMAX subscriber stations. Though, as we mentioned previously, the combination of a WiMAX based backbone for Wi-Fi mesh networks seems to be an attractive option. Moreover, mobile WiMAX will be definitely the solution when mobility is of key importance.

Cellular Mobile

3G cellular systems together with enhancements like HSDPA/HSUPA (High-Speed Downlink Packet Access/High-Speed Uplink Packet Access) (ITU, 2008) also, due to the smaller cell size, offer per-customer data rates that would satisfy the requirements of most of today's mobile applications. Nevertheless, it is hard to find community networks that are based on cellular mobile service. The reason might be a simple one: municipalities did not take this option into account. On the other hand, cellular operators might be also reluctant to work out individual offers for cities with very special pricing, and specific solutions in addition to cellular coverage, e.g. a combination with WiMAX, to support large institutional users. Thus, we include this option here for completeness only.

Technology Selection for Wireless Community Networks

As mentioned previously, there are differences between community networks and telcos' networks from the viewpoint of technology selection, too. For community networks, cost minimization is not the primary objective and implementers of CNs can also experiment with new and advanced technologies. Another difference is that interoperability is of critical importance for CNs.

There are several factors to be considered when selecting the right wireless technology for a CN, such as application requirements, coverage, timeframe, frequency issues and costs. Below we focus only on the requirements of the applications and expected coverage which should be identified in the first step. A summary of our analysis about the state-of-the-art wireless technologies is shown in Tables 1 and 2. One of the most important issues in technology selection is finding the most suitable alternative for the application requirements. Table 1 helps choose the right technology and configuration by coverage, bandwidth and density parameters. Table 2 focuses on QoS (Quality of Service) measures.

First, we should define the main elements of the network:

- Microcell is an area covered by one access point or mesh node in the access network.
- Macrocell is a union of well-connected microcells. Macrocell connects to the backbone with one or more backbone access points.
- Backbone access point (BAP) is a node of the network, which connects one macrocell to the backbone network.

As a general assumption, there is no sectorization in the discussed network topology scenarios, we use only omni-directional antennas.

Explanations to Table 1 column by column:

1. ID: Identifies the cases for referring.
2. Technology: What network technology, Wi-Fi or WiMAX, is used and in point-to-multipoint or in mesh mode.

Table 1. Technology selection for capacity and coverage planning

ID[1]	*Technology* [2]	*Configuration* [3]	*Maximum microcell capacity* [4]	*Number of microcells in macrocell* [5]	*Maximum macrocell capacity* [6]	*Maximum microcell radius* [7]	*Maximum node distance* [8]	*Maximum coverage (macrocell size in 0.01 km²)* [9]	*Maximum bandwidth density (Mbps/0.01 km²)* [10]
1	Wi-Fi	NLOS	20 Mbps	1	20 Mbps	100 m	160 m	3	7
2	Wi-Fi mesh	Max. 2 hops NLOS	7 Mbps	25	175 Mbps	100 m	150 m	50	3.5
3	Wi-Fi mesh	Max. 3 hops NLOS	2 Mbps	85	170 Mbps	100 m	140 m	150	1
4	WiMAX	LOS	100 Mbps	1	100 Mbps	3 km	3 km	1000	0.1
5	WiMAX	NLOS	50 Mbps	1	50 Mbps	1 km	1 km	100	0.5
6	WiMAX mesh	Max. 2 hops NLOS	16 Mbps	25	380 Mbps	1 km	1 km	2500	0.15

3. Configuration: Indicates the condition of propagation, which affects link quality and cell size, and the maximum number of hops in mesh scenarios.
4. Maximum microcell capacity: Average effective usable data rate at network layer using the given technology and configuration. 100 Mbps at WiMAX is based on more than 20 MHz as bandwidth.
5. Number of microcells in a macrocell: This number is 1 in non-mesh networks. In mesh networks, it can be calculated from the maximum number of hops.
6. Maximum macrocell capacity: Microcell capacity multiplied by the number of microcells in a macrocell.
7. Maximum microcell radius: These parameter values are based on transmission power limited by EU-conform regulations at high data transfer rates for high cell efficiency. Within the given range, the actual technology can provide an almost perfect coverage in most cases.
8. Maximum node distance: The distance of two neighboring nodes should not be longer than the maximum microcell radius multiplied by the square root of 2, caused by the square gridded layout of the topology. Hence, the full coverage is ensured with existing overlapping areas.
9. Maximum coverage (macrocell size in 0.01 km^2): This value shows the size of the covered area by one macrocell. We can also define backbone access point (BAP) density in number of BAP/km^2 which can help estimate the initial and operational costs.
10. Maximum bandwidth density (Mbps/0.01 km^2): The ratio of the capacity and the coverage of macrocell. It shows the available bandwidth on each area of 100x100 m^2.

The data in Table 2 are also based on our measurements and calculations using the following assumptions:

- Soft QoS means IEEE 802.11e standard in Wi-Fi. Managing QoS is one of the inherent features in WiMAX.
- The delay and jitter parameters are one-way latency measures.

Our practical suggestions for technology and configuration selection based on Table 1 and 2 are as follows:

- If some not frequently connected spots should be covered by a wireless network,

Table 2. Technology Selection for QoS Planning

ID	*Technology*	*Configuration*	*Maximum microcell capacity*	*Average delay per hop (low utilization)*	*Average delay per hop (high utilization) without QoS*	*Average delay per hop (high utilization) with QoS*	*Band-width allocation capability*	*Voice transmission capability with/without soft QoS*
1	Wi-Fi	NLOS	20 Mbps	5 ms	400 ms	100 ms	no	yes / no
2	Wi-Fi mesh	Max. 2 hops NLOS	7 Mbps	10 ms	1000 ms	200 ms	no	yes / no
3	Wi-Fi mesh	Max. 3 hops NLOS	2 Mbps	25 ms	2000 ms	400 ms	no	no / no
4	WiMAX	LOS	100 Mbps	20 ms	100 ms	50 ms	yes	yes
5	WiMAX	NLOS	50 Mbps	30 ms	150 ms	50 ms	yes	yes
6	WiMAX mesh	Max. 2 hops NLOS	16 Mbps	80 ms	300 ms	100 ms	yes	yes

standalone Wi-Fi access points as hotspots should be used. It can be used in LOS and, to a limited extent, in NLOS conditions. IEEE 802.11e capable devices should be used to support QoS requirements to real-time services such as voice communication (Table 1, 1st row).

- If a larger area has to be covered by a limited number of backbone access points, Wi-Fi mesh network with only few hops should be used. More than 2-3 hops to the BAP cause degradation in effective bandwidth and in QoS parameters, too. Real-time applications can tolerate this relapse up to 2 or 3 hops with 802.11e support (Table 1, 2nd and 3rd rows).
- Wide areas with low density of users should be covered by WiMAX. It can be used not only in access networks but also in backbone networks in point-to-point or point-to-multipoint configuration. Robustness and high data rates of WiMAX guarantee the QoS and sufficient capacity in LOS and in NLOS environment, too (Table 1, 4th and 5th rows).
- WiMAX can operate also in mesh mode. In this case, advantages of Wi-Fi mesh and WiMAX are combined. However, this solution has been not widely implemented yet (Table 1, 6th row).

To summarize the possibilities for technology selection to build wireless CNs we can say that, for a number of applications, Wi-Fi mesh could be the solution, but for applications that require QoS and high bandwidth, WiMAX is the best choice. However, because of the low penetration of WiMAX devices, we have to use still today a widely preferred access technology, such as Wi-Fi. On the other hand, the backbone or distribution network should be robust and should have sufficient capacity. The combination of WiMAX and Wi-Fi technology, and the combination of mesh, ordinary access and transfer can be the optimal solution for every wireless community network. Wi-Fi will remain the only feasible customer access solution for the next 2-3 years (until mobile WiMAX cards will be as ubiquitous and cheap as expected by major market players).

DESIGN OF WIRELESS COMMUNITY NETWORKS

In this section, we deal with issues related to the design of wireless community networks. First, we overview the design methodology, then illustrate the network design by a case study.

Design Methodology

In general, there are significant differences between planning of CNs and ISPs' or other service providers' design methodology. Key differences include the following requirements for the planning of CNs: (i) ubiquitous Wi-Fi access covering the whole territory of the community; (ii) users should be provided with other forms of access as well; (iii) mobility or at least nomadic access across the covered area must be supported; (iv) support of a multiplicity of user devices from simple mobile phones through PDAs and laptops to video conferencing equipment; (v) the network should support a specific set of government, business and society-related applications.

The whole design process consists of the following steps:

1. Identifying applications and services. First, we should select the key applications and services which raise requirements toward the network.
2. Identifying network technology requirements, based on applications. We should analyze the requirements of the applications and services selected in the first step. This analysis should contain QoS (delay, jitter) and bandwidth parameters.

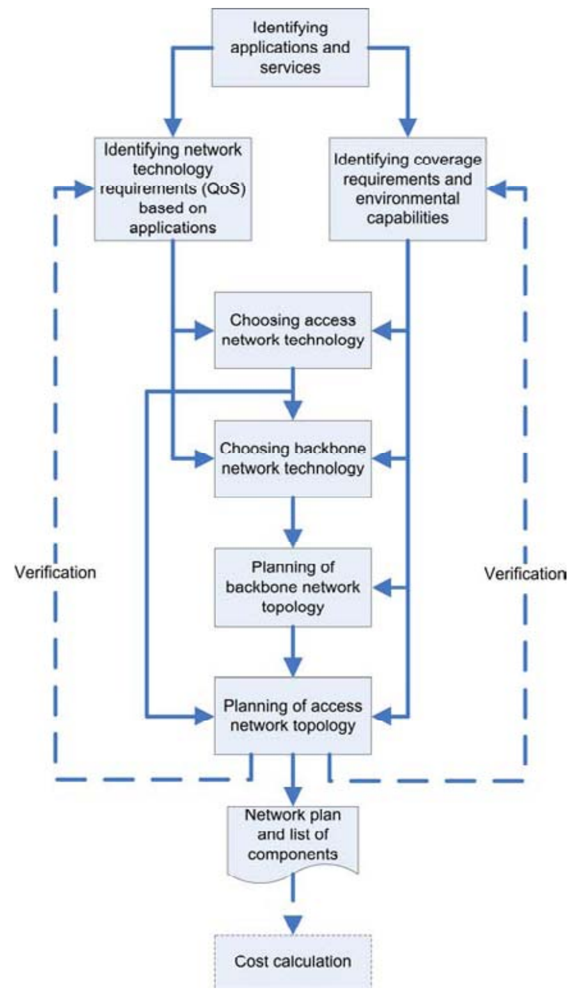
3. Identifying coverage requirements and the possibilities and limitations of the environment. To prepare the network technology selection, we should determine the area which is supposed to be covered by the network, with its topography, natural obstacles such as hills or trees as well as buildings, availability of support structures, towers etc.
4. Choosing network technology. Selecting the right technology is one of the key parts of network planning. This decision should be based on identified requirements and conditions of the environment. We should choose optimal solutions both for the access and the backbone network. This step of the design process is explained in detail in Section 3.2.
5. Planning of network topology. This complex part of the methodology uses the results of the coverage requirement analysis as well as the network technology selection. We should plan the network topology according to the topography and the optimal station placement strategies.
6. Verifying original requirements. Last, but not least, this step stands for verifying the results of planning. We should recognize the differences between the original requirements and the capabilities provided by the planned network.

These steps are illustrated in Figure 1.

Case Study: Network Design for a Digital City in Hungary

To illustrate the design methodology discussed previously we present here the wireless network design of a digital city in Hungary. The digital city project is the municipality's initiative to implement a city-wide network infrastructure and services based on that.

Figure 1. Flowchart of the design process for wireless community networks



Services for the Municipality of the City

The planned wireless infrastructure will serve several important goals: (i) it will carry the internal data and voice traffic among public institutions and publicly controlled companies, thus saving costs of bills currently being paid to telecom service providers; (ii) it will improve the efficiency of work processes and introduce electronic customer services via an e-government initiative; (iii) it will improve services for citizens and facilitate citizens' participation in public processes. Some specific applications based on interviews with potential large users are as follows:

A) Public safety system

The objective is to improve public safety and reducing crime in the city by establishing a network of surveillance cameras and equip police and fire brigade personnel with wireless enabled devices.

B) Telemetrics for a Local Utility Company

The objective is to use the wireless network to implement AMR for the local water company.

C) Parking Management for a Local Parking Company

The wireless community network is planned to support parking services in several ways: communication with the parking ticket dispensers, providing enforcement staff with handheld devices, etc. Additional services include online payment, SMS warnings of expiring parking tickets and reminders of unpaid parking dues.

D) Services for a Public Bus Company

The objective is to improve the efficiency of the company's operations and the quality of passenger service. The planned wireless CN will collect and transmit real-time data related to departure and arrival times, delays, technical problems, road and traffic conditions.

E) Advanced Tourism Information System

The objective is to implement a tourist and cultural information portal based on geospatial information system, and install several kiosks supporting free or low-cost Internet access to this portal.

Pilot Network Estimations and Planning

After identifying the services we have to analyze their requirements (Table 3). The required overall microcell capacity is calculated as the aggregate of average bandwidth for each service, and the maximum value of delay in the network must be not greater than the minimum of the maximum tolerated service delays.

We have studied two different scenarios:

Scenario 1: Fixed WiMAX as backbone network and Wi-Fi access from WiMAX subscriber stations. In this scenario, the user connects to a Wi-Fi access point, which is connected to a WiMAX BAP through WiMAX by its secondary interface. WiMAX BAPs connected to each other are the main points of the backbone network.

Scenario 2: Wi-Fi mesh with WiMAX backbone and interconnection network. In this scenario, the user connects to a Wi-Fi mesh access point, which can forward the traffic to another one in mesh mode. Finally, the traffic reaches one of the mesh nodes which

Table 3. Requirements of the identified services

Service ID	Bandwidth per user per microcell	Probability of activity per user	Average number of user per microcell	Average bandwidth per microcell	Maximum tolerated delay
A	1 Mbps	1	2	2 Mbps	500 ms
B	1 kbps	<0.0001	50	< 1 kbps	2000 ms
C	10 kbps	<0.001	50	< 1 kbps	500 ms
D	10 kbps	0.1	5	< 5 kbps	1000 ms
E	1 Mbps	0.2	5	1 Mbps	500 ms
Minimum required capacity and delay limit:				3 Mbps	500 ms

Table 4. Comparison of Scenario 1 and 2

Scenario	1	2
Number of WiMAX nodes in backbone	8	2
Number of Wi-Fi access points with a secondary WiMAX interface	42	0
Number of Wi-Fi mesh nodes	0	36
Number of Wi-Fi mesh nodes with secondary WiMAX interface	0	6
Average cell capacity	15 Mbps	5 Mbps
Overall capacity of the network	400 Mbps	200 Mbps
Worst case estimated delay	250 ms	300 ms
Worst case estimated jitter	100 ms	200 ms

has a WiMAX interface to connect to the main WiMAX base station.

Regarding these scenarios Table 4 contains the numerical results calculated by using the previous considerations and tables. Unfortunately, due to space limitations we cannot present the detailed planning process here.

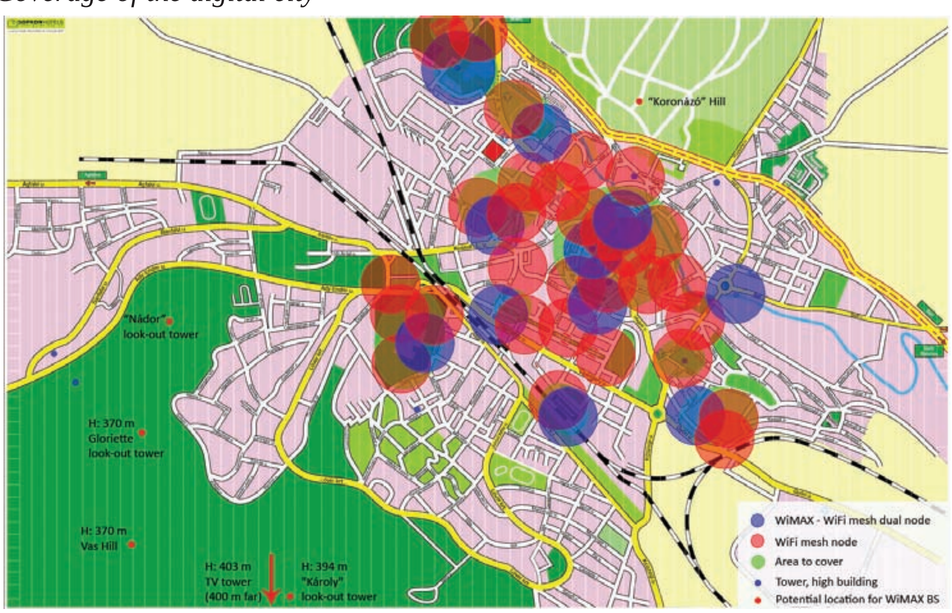
If we consider only the technical parameters Scenario 1 is a better choice, but it uses plenty of Wi-Fi access points with WiMAX interface. Therefore if one should keep the costs low Scenario 2 is preferable.

Map Based Planning

After the general planning phase the real map based planning follows. First, we should identify the type of places, which are to be covered, such as the old town, tourist destinations, educational institutions or main routes of public transportation. The wireless nodes should be placed mainly at buildings owned or controlled by the city like educational institutions, governmental buildings, etc.

Then, we should place the nodes to the potential locations covering the selected areas. The ratio of overlapping area can exceed a predefined limit only in the old town and some other places, where the density of wireless users is higher than the average.

Figure 2. Coverage of the digital city



We should select some high points of the city, e.g. churches, towers (blue dots on Figure 2), where the nodes with dual (WiMAX and Wi-Fi mesh) interfaces (blue circles on Figure 2) can be placed. A node with dual interfaces aggregates the traffic from its cell and the traffic of 3 other Wi-Fi mesh nodes (red circles on Figure 2) on average, and forwards it to the backbone access points using WiMAX. The number of dual nodes should be equal or higher than the number of nodes in the previous estimation. More dual nodes can rise capacity and fault tolerance in the network, but also additional costs are incurred. Therefore we decided for additional dual nodes but keeping the total number of cells close to the predicted values. Due to capacity and distance aspects 2 BAPs are enough to serve the whole area to be covered, but we should consider 3 BAPs to ensure redundancy and avoid blind spots. Fortunately, the city is surrounded by hills from southwest and from north. Hence, several look-out towers exist near the city as potential location of BAPs. These spots are marked with red dots on the map (see Figure 2).

BUSINESS PLANNING

In this section, we deal with issues related to business planning. First we discuss the general considerations for business planning, implementation, and operation and maintenance of community networks. Then we describe a common business structure representing a public-private cooperation. And finally, we discuss the most important aspects of business model creation.

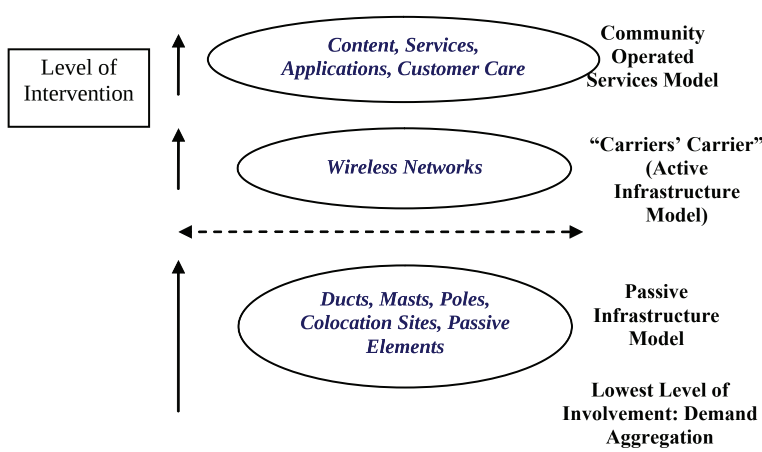
General Considerations

The interesting feature of business models for a public entity is that getting the invested money back in short term is not of primary importance. Thus, longer ROIs (Return On Investment) are acceptable, and maximizing the profit is not the primary objective as, there are important indirect benefits which result from aiding new service providers, ISPs, telecom companies, value added service providers to enter the market and grow. Hence, the public entity can obtain additional revenues from the company taxes. Second, the public sector can significantly decrease its expenses for telecom services using the public entity's own infrastructure.

There is no an easy and straightforward business model to deploy and operate community networks. In the following, we discuss some possibilities according to the involvement of the public entity.

Figure 3 shows three basic models and a fourth one called "Demand Aggregation" which

Figure 3. Basic models of public involvement



represents the lowest possible level of public participation (Chlamtac et al., 2005). The first basic model refers to a low level of public involvement. In this case the public entity acts as only a passive infrastructure provider. The highest level of public participation is when the public entity acts as a service provider. This model should be applied carefully since it creates a conflict of interest situation. The local government may provide services only for internal purposes, i.e. for public institutions, thus avoiding competition with service providers in the marketplace. The model in between is a pure wholesale model when the community network operator acts as "Carriers' Carrier".

The aforementioned models differ in terms of ROI, too. Figure 4 illustrates the approximate ROI values for different levels of public involvement (Chlamtac et al., 2005). The public entity should expect the longest ROI value (10 – 15 years) if the passive infrastructure model is applied. A bit shorter ROI period (3 – 5 years) can be assumed when the active infrastructure model is used. And the shortest ROI value (2 – 3 years) can be considered when the public entity acts as a service provider.

The participation of a public entity in creating and operating a community network is often accomplished in a kind of public-private partnership (PPP). The typical models according to the structure of public-private cooperation are as follows (Chlamtac et al., 2005):

1. Publicly owned and operated
2. Privately owned and operated
3. Non-profit owned and operated
4. Publicly owned, privately operated
5. Owned and operated by a public utility
6. Privately owned and operated jointly with the municipality

The choice of the appropriate model is also influenced by regulations that may allow or restrict the different ways and levels of how a public entity can participate in providing telecommunication services. Moreover, the selection among the possible models can be based on costs and/or complexity of management for the public entity. In Figure 5, the different models are arranged in a cost/complexity coordinate system. The "publicly owned and operated" model has the highest value of both the cost and complexity levels from the viewpoint of the public entity. At the other end of the spectrum the "privately owned and operated" model is located, while in between the other combinations offer different cost and complexity levels.

Business Structure

The business structure can be based on the following strategic alternatives: (i) building the community network by the LG (Local Government); (ii) teaming up with a local company and build the community network together. In the

Figure 4. ROI for the Basic Models of Public Involvement

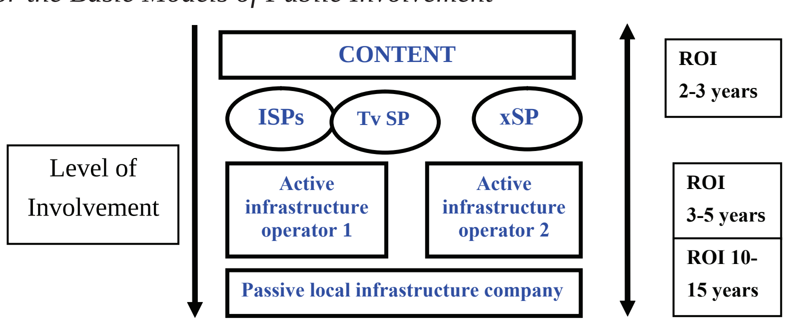
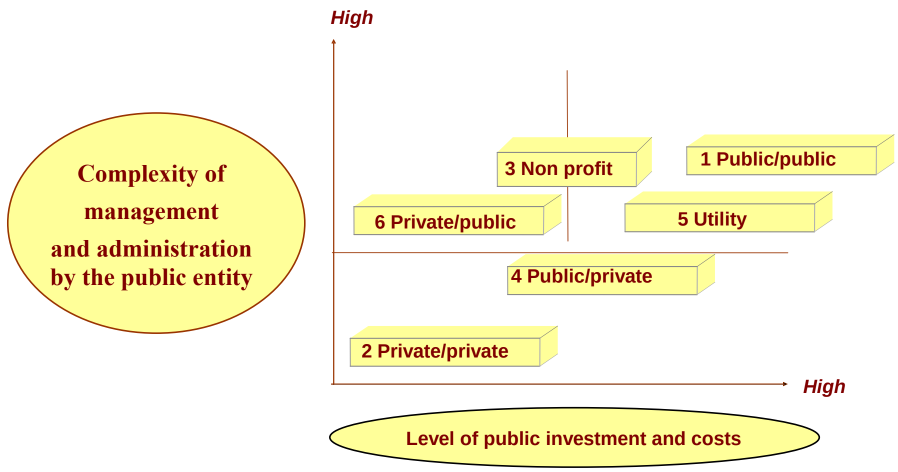


Figure 5. Comparison of Public-Private Partnership Models



following, we analyze the second option as an "incarnation" of a public-private cooperation. We consider the following business structure (Szabó et. al, 2004):

a. *Infrastructure* company: set up by the LG and the selected local company. It makes the investment in the wireless infrastructure, owns the wireless network and offers the use of the native wireless infrastructure as a product to the internal customers within the community and to the external market. Internal needs and requirements are channeled through the Services company.
b. *Services* company: is to be set up to take care of the internal services, including Internet access, voice, data and others, for the public sector. The Services company operates on business terms but does not sell its services in the open market.

The ownership structure (see Figure 6) is that the Services company is majority-owned by the Infrastructure company, but it can also have minority interests from partners. On the other hand, the majority owner of the Infrastructure company is the local government which can team up with some other private partners as minority owners.

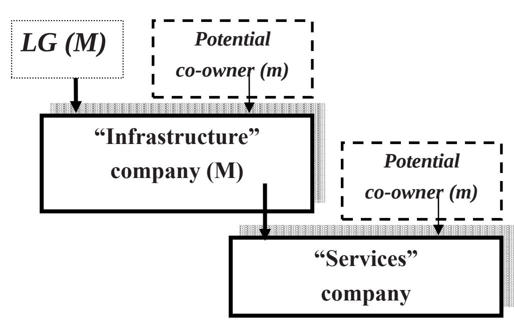
Creating the Business Model

The most important part of building a business model is to work out the set of assumptions for the calculations. Here we discuss only these assumptions which can be summarized as follows.

Voice traffic, internal needs: It is important to assess the fraction of internal voice traffic on the community network. Typically this value is about 40% – 50% and that a 20% – 25% cost cut can be achieved on the external traffic by using the community network before reaching the external world.

Data traffic, internal needs: On the data and information related traffic, it is reasonable to assume a conservative scenario, where the

Figure 6. Ownership structure



bandwidth growth and the price reductions will compensate each other.

Wholesale of excess capacity: Scenarios are to be drawn up for two different types of customers: (i) telecom operators; and (ii) business customers. It can be usually assumed that 1-2 out of the potential telecom operators and ISPs will be future customers of the community network.

Investments: Here we need a total investment figure, the total length of the investment period and the division of investments over that period. The model should also include depreciation calculation for different periods.

Financial assumptions: Include the total needed equity and its division into own equity and external financing. The repayment conditions of external financing have to be taken into account, too.

Operational costs: The operational costs can be split into three main groups:

- Operation & Maintenance for the network. It includes external service contracts and costs for technical personnel;
- Personnel costs. Its estimate (excluding technical staff) can be based on management experience;
- Other costs, including administration, marketing, etc. All these costs can be calculated as a percentage of the revenue which gives a fairly precise figure.

FUTURE TRENDS

The area we dealt with in this chapter is a rapidly evolving one and we can only draw some development trends and their possible outcomes with a degree of uncertainty. Let's do it by addressing the main aspects of our integrated treatment: *applications, technologies* and *deployment models.*

A. Applications

In this field, the most significant trend is the development of new applications based on geospatial information systems. The success of Google Earth and the increasing number of web-based applications that rely on its geo-information services is an important example of how geodata have become pervasive in our society. Communities are geographic and geospatial entities. Workers in those communities have come to rely on geographic information in order to perform their jobs. Such information is vital for making decisions in a variety of areas including crime management, business development, flood mitigation, environmental restoration, community land use assessments and disaster recovery. Advanced GIS - Geographic Information Systems have become key tools for both local governments, utility providers, and a variety of other public and private sector organizations. Moreover, citizens and tourists alike want to be able to access maps, get directions, find shops, restaurants and lodging, and learn about local attractions and programs, all from their mobile device. The underlying location technology is the GPS (Global Positioning System), well known in everyday life. A recent alternative to the satellite-based GPS is based on the network of Wi-Fi access points. Skyhook's Wi-Fi Positioning System (WPS) (Skyhook, 2008) is the world's first location platform to use the Wi-Fi capability already present on a mobile device to deliver positioning. The service is currently available across the US and being extended to other parts of the world. One of the advantages of this technology is that it provides a high degree of indoor availability.

B. Wireless Technologies

The work on standardization of IEEE 802.16 (WiMAX), including mobile WiMAX, has already been completed. However, its implementation is left now to industry. Mobile WiMAX is not yet ready for use but since there are industry giants, such as Intel, behind it is very likely that it will become reality soon. There are competing technologies including recent enhancements of the 3rd generation cellular mobile systems and the emerging 4G or, more exactly, B3G (Beyond 3G) systems. We should also mention the IEEE 802.20 standardization effort, also known as MBWA (Mobile Broadband Wireless Access) where the work is in progress yet and if everything goes well the result may be a very attractive, wide area, broadband mobile technology solution that could become a real competitor to 3G and B3G systems.

It is hard to predict which of the aforementioned emerging technologies will win. Most likely, a co-existence and co-operation among them will lead to the next generation broadband wireless/mobile systems. All of these emerging technologies will have hard time to compete not only with each other, but, most importantly, with the IEEE 802.11 based WLAN systems, as the latter have recently been significantly improved in terms of data rates, area coverage due to the mesh extension and quality of service provisioning capabilities.

C. Deployment Models

There is a large variety of available business models, and new ones or innovative combinations will emerge. Let us point out to a particularly interesting emerging model which relies on citizens' participation in creating and operation of the network.

We know that at the beginning of the development of community network movement, there were many grass-roots initiatives, while at present most of the CN projects are publicly-driven. However, it looks like the community participation takes a new form in the approach followed by FON, a global Wi-Fi bandwidth sharing community (FON, 2008). FON is the largest Wi-Fi community in the world. Its mission is to stimulate the growth of Wi-Fi Internet access around the world by creating a global community of "Foneros" (consumers who agree to share their Wi-Fi in return for free access to all other Wi-Fi access points in the Community) and offering low-cost access to non-community members (called "Aliens"). FON's business model is based on the money that is generated from Aliens purchasing FON passes. FON is collaborating with various cities, municipalities, and communities all over the world. It also has a number of cooperation agreements with ISPs in several countries, among them some of the big ones such as British Telecom in the UK.

CONCLUSION

In this chapter, we have presented an integrated view of applications, technologies and business models for wireless community networks. We have discussed the technological aspects and proposed a design methodology illustrated by the example of an ongoing digital city project in Hungary. Then we have analyzed the relevant business models for community networks. Before concluding our integrated presentation we have also outlined some future trends related to the main aspects dealt with in this chapter. And finally, it is worth mentioning that there are still room for further research regarding novel wireless technologies and design methods for existing ones such as Wi-Fi mesh and WiMAX (including mobile WiMAX) networks.

REFERENCES

Chlamtac, I., Gumaste, & Szabó, Cs. (Eds.). (2005). *Broadband Services: Business Models and Technologies for* Broadband Community *Networks.* Wiley.

Corpus Christi Pioneers Metro-Wide Wi-Fi Mesh Network for AMR. (2007, July). *A Tropos Network Case Study*.

Edwards, P. N. et al. (2007, January). Understanding Infrastructure: Dynamics, Tensions, and Design. Report of a Workshop on *History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures*. NSF.

FON. (2008). Retrieved January 30, 2008, from http://www.fon.com/

IEEE 802.11. (2008). *IEEE 802.11 Standard Family*. Retrieved January 30, 2008, from http://standards.ieee.org/getieee802/802.11.html

IEEE 802.16. (2008). *IEEE 802.16 Standard Family*. Retrieved January 30, 2008, from http://standards.ieee.org/getieee802/802.16.html

Intel. (2005, August). *Core Technologies for Developing a Digital Community Framework Wireless Fabric, GIS Portal, and Government Federated Services Bus Technologies* (White Paper)

Intel Solutions. (2005). *Digital Community Best Practices* (White Paper).

ITU (2008). *3G network standards*. Retrieved January 30, 2008, from http://www.itu.int/ITU-T/

Longano, S. (2007, May). *T.Net – Trentino in Rete.* Paper presented at the "Bridging the Broadband Gap" conference, Brussels.

MuniWireless. (2008). Retrieved January 30, 2008, from www.muniwireless.com

Niyato et al. (2007, February). IEEE 802.16/WiMAX-based Broadband Wireless Access and its Application for Telemedicine/e/Health services. *IEEE Wireless Communications*, pp. 72-83.

OPAALS. (2008). *Open Philosophies for Associative Autopoietic digitaL ecosystemS*. Retrieved January 30, 2008, from http://www.opaals.org

Settles, C. (2007, June). Making the Case for AMI and ROI: Taking the automated meter reading (AMR) to the next great level. *Municipal Wireless Snapshot Report*.

Skyhook. (2008). *Wi-Fi Positioning*. Retrieved January 30, 2008, from http://www.skyhookwireless.com/

Star, S.L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructures: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1):111-134.

Szabó, Cs., Chlamtac, I., & Bedő, E. (2004, January). Design Considerations of Broadband Community Networks. In *Proceedings of the Thirty-Seventh Annual Hawaii International Conference on System Science (HICSS 2004)*. Hawaii, USA: Computer Society Press

Szabó, Cs., Farkas, K., & Horváth, Z. (2008). Motivations, Design and Business Models of Wireless Community Networks. To be published in *Mobile Networks and Applications* journal: Springer

Szabó, Cs., Horváth, Z., & Farkas, K. (2007, October). Wireless Community Networks: Motivations, Design and Business Models. In *Proceedings of the Third Annual International Wireless Internet Conference (WICON 2007)*. Austin, Texas, USA.

The Wireless Philadelphia Executive Committee. (2005, February). *Wireless Philadelphia Business Plan: Wireless Broadband as the Foundation for a Digital City*.

Webb, W. (Ed.) (2007). *Wireless Communications: The Future*: Wiley.

KEY TERMS

Business Model: Describes a broad range of informal and formal models that are used by enterprises to represent various aspects of business, including its purpose, offerings, strategies, infrastructure, organizational structures, trading practices and operational processes and policies.

Community Network: The combination of the telecommunication infrastructure, the services provided upon it and the specific business model to operate the infrastructure and provide services.

Digital Divide: The gap between people with effective access to digital and information technology, and those without access to it.

Digital City: Initiative of the municipality to implement a city-wide network infrastructure and public/private services upon it.

PPP: Public-Private Partnership, a form of business co-operation between public and private entities.

ROI: Return On Investment, the period in time until the invested money returns back.

Wi-Fi Mesh: Wireless-Fidelity mesh, a peer-to-peer multi-hop network topology based on the IEEE 802.11 standard family, where the nodes cooperate with each other to route information packets through the network.

WiMAX: Worldwide Interoperability for Microwave Access, an emerging wireless technology and a flexible telecommunication architecture based on the family of IEEE 802.16 standards.

Chapter XLII
Human Factors for Business Mobile Systems

David Golightly
University of Nottingham, UK

ABSTRACT

Technology has to be fit for purpose in order to deliver its promised business impact. An important component of this is making sure technology offers its intended users the functionality they require in a manner appropriate to how they want to achieve their goals. This chapter presents an area of science, human factors, with particular relevance to developing and deploying business mobile systems that are fit for purpose. The value of human factors is discussed, before moving onto some general guidelines to ensure that a mobile technology is suitable for its users. Finally, each system, each application, and each user should be taken on their own merits. To achieve this, it is critical to take an approach that puts users at the heart of requirements, build and deployment of a new technology or service. This approach is presented as a three-stage model of context analysis, specification and design, and evaluation.

INTRODUCTION

Mobile technology is now delivering new opportunities in the business arena, with two critical characteristics

- **Portability:** The size of many mobile devices allows them to be taken to, and used in, an almost infinite range of locations and situations.
- **Connectivity and communication:** Typically, these devices are primarily for communication (e.g. the mobile/cell phone or Blackberry-style email client). This is being combined with increasing processing power, and therefore increasing functionality of applications on the device, to make rich and complex user experiences.

We are probably only beginning to scratch the surface of how such devices can be used to support business tasks, but this power comes with limiting factors. The portability of devices means they have to be small, yet support a complex range of user inputs and device outputs; their versatility means that interaction must be possible in environments that impinge on the user's experience much more than the relatively calm and constrained surroundings of the average office, and the new complexity of applications must be presented in a way that hides the complex interplay of device resource sharing, switching between applications, and changes in network connectivity that is occurring behind the scenes. Most important of all, mobility demands that the technology is fit for purpose within the context of use, to the point where aspects of its functionality should be made invisible for a given context (Bergman and Norman, 2000; Pascoe, Ryan and Morse, 2000).

In the following, the field of human factors, and its contribution to developing fit for purpose technology, is presented. This is followed by guidance on how a human factor approach can be applied to the design, development and management of business telecommunications – specifically mobile business applications. This guidance is presented in two parts. The first part covers key human factors considerations for effective mobile business applications. The second part, presents the steps to follow within an iterative user-centred design process, to make sure the mobile business application meets the needs and abilities of potential users. These two parts are complementary in that key human factors considerations should be considered within the development process. This guidance is summarised in Figure 1.

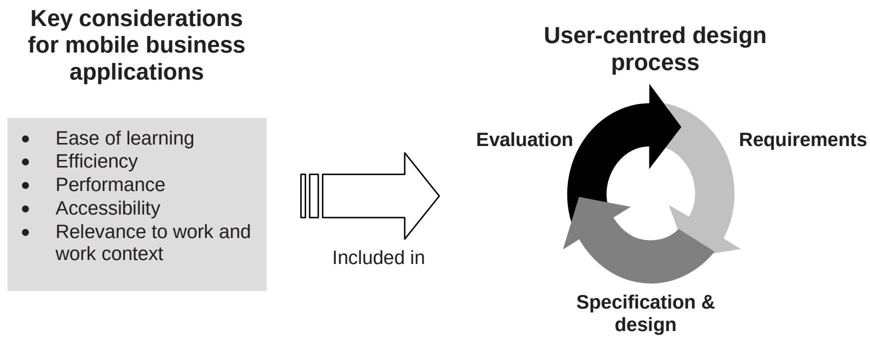
Future directions for the field and concluding comments are then presented.

BACKGROUND

Combining the capabilities and constraints of mobile technology in an effective manner makes for a seamless experience and a powerful work tool. This quality is commonly referred to as usability – "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."(ISO 9241-11, 1998). Usability ensures that people will have the objective outcomes of greater productivity, better efficiency, and fewer errors, while having the subjective experience of greater confidence and satisfaction in terms of achieving set goals, and a greater propensity to re-use a product.

Achieving this is not a trivial task. All too often, mobile projects focus on getting the technology to work as a technical challenge, but less on ensuring that it is designed to fit with how users operate. This may offer a significant risk to a project. At worst users will completely reject mobile device or service (either as a whole, or as a specific ap-

Figure 1. Human factors guidance for mobile business applications



plication within a multi-function device) and that project will effectively fail. Even in better cases, users will use the technology but only through using work-arounds or by making repeated errors, which severely impact performance and productivity. Also, because deployed systems are either often built, as bespoke, by a third party or bought off-the-shelf from a supplier, understanding user needs is just as relevant to procurement decisions as it is to design decisions (Forman et al, 2007).

A remedy is to take a human factors approach to overcoming these obstacles. Human factors is a domain offering research and guidance on how to design work, and the tools of work, to provide the best match with work goals, people's needs, and human physical and cognitive capabilities (Wickens et al, 2006). The goals of human factors are to enhance performance, increase safety, and increase user satisfaction for human interaction with any system – whether that is a technical system (such as a mobile device) or a socio-technical system – an integration of technology, communication, procedures and team work (Wilson, 2000).

Human factors consists of a body of knowledge on human capabilities. In physical terms this covers areas such as anthropometrics, physiology, and the impact of manual activity on the body (this is often presented under the label of 'Ergonomics'). In environmental terms, this covers the effects of excessive cold, heat, noise and other factors that influence safety comfort and productivity. Finally, psychological factors are considered – these are most often cognitive, such as the capacity of attention, the characteristics of memory, problem solving and learning, and the nature of sensory apparatus. For a comprehensive overview of the field of human factors, refer to Wickens et al, 2006.

As well as the body of knowledge on capabilities, human factors offers a suite of tools for understanding and evaluating work contexts (Wilson and Corlett, 2005), and integrated methodologies for applying such tools to a workplace or system development (see Garrett, 2003, for a design-oriented approach; see Pew and Mavor, 2007, for a more technical approach, considering issues such as risk analysis). This approach of applying an understanding of human needs, capabilities and design is often referred to as a *user-centred* approach (see Vaananen-Vainio-Mattila and Rutuska (2000) for description of how a user-centred design approach is applied at Nokia).

Human factors has specific value to business mobile applications. First, it is critical that any device is designed to make best use of human needs and capabilities, and human factors offers substantial guidance on this area. Second, taking the previously stated ISO description, usability has to be determined for a 'specific context of use'. For a mobile business application that means a work context. All too often, technology is developed in isolation without an appreciation of the context of use, and the resulting constraints that impinge on technology use. Taking a human factors approach can ensure that the larger work context is understood, and reflected in the subsequent product or service.

The rest of this chapter will use as its focus mobile business applications – functionality that might be available natively, or through network connectivity, on a dedicated handheld device. That said, the guidance and approach proposed is applicable to all forms of application in the telecoms arena (e.g. support, Gray et al, 1993). Also, guidance on how to design a mobile business application is just as relevant when deciding which application to procure. Issues such as acceptable system response may be particularly relevant when setting Service Level Agreements (SLAs) with a service provider. Finally, while B2C products and services are not discussed directly, much of the research, guidance and approaches covered are just as relevant to the consumer domain (see Laukkanen (2007) for an example of a specific evaluation of mobile services aimed at consumers).

CHARACTERISTICS OF A USER-CENTRED MOBILE APPLICATION

The following section presents a number of considerations or characteristics to be borne in mind to deliver a user-centred mobile business product or service. For simplicity, the term *mobile business application* will be used. This means any business-related function or application that can be delivered through a mobile device. In reality, mobile devices tend to include a number of applications that may be used in combination, but the points presented here should still be applicable even if functions are used in combination. This list is not exhaustive. In particular, physical ergonomics of the handset including input and output mechanisms have not been covered directly. For more information on these issues refer to Brewster (2002), Weiss (2002), and Jones and Marsden (2006). For more fundamental coverage of the human factors surrounding the vision and auditory system (affecting text recognition, alarm discrimination and the comprehension of speech of a limited acoustic spectrum) refer to Wickens et al (2006).

Ease of Learning

An application that is easy to learn will require less training, less support and less time until it plays a productive role in the user's work. As important as supporting the complete novice, ease of learning supports those users who may only use an application infrequently.

Norman (1988) lists a number of characteristics that make a device easy to learn.

- **Visibility:** Whereby the user can tell the state of application, and also quickly see what functionality is available, with the most important functions being most prominent, as opposed to ordered alphabetically, arbitrarily or by which service provides the highest revenue to an operator. What is important may vary from work context to work context and must be carefully understood before making a design decision. This will be informed within a user-centred design process by a complete 'context analysis' – discussed later. Visibility does not have to be literal – the structure of interaction should be made clear, even when working with audio-based menus (Howell et al, 2006).
- **A conceptual model:** Users will try to build a model of how the various elements of a device or application fit together so they can make assumptions about how it will work and predict either its actions, or responses to user inputs. The application should therefore present a coherent model of how all the aspects of the functionality work together, so the user is informed by accurate information, rather than guesswork. Conversely, only those aspects of the system that the user needs to actively engage with should be salient to the user (for instance, it may not be necessary to repeatedly ask the user whether they want to save data to the sim versus a separate memory card). The power of a strong conceptual model has been demonstrated repeatedly to enhance learning of a device (Payne, 2003). Consistency is also critical to this. The application should be internally consistent (similar names, icons etc mean similar things and do similar things), but should also be consistent with other applications on the same device, and with other technology and work terminology.
- **Good mappings:** There should be a clear relationship between actions and their effects, so that it is clear what the outcome will be once the use makes a choice. This allows the user to guess an outcome with certainty rather than making an error or having to refer to some form of support. Jones and Marsden (2006) give the example of the iPod wheel as

an example of a strong mapping, where the clockwise/anticlockwise operation clearly maps to more/less or louder/softer.

- **Feedback:** The application should provide information about itself in a timely manner (see 'Performance') and in a manner that makes sense to the user. Error messages and dialogs should be clear and concise, and, particularly for mobile devices, system status such as battery level and network connectivity must be clearly displayed.

Efficiency

While ease of use can get the user started with their interaction, it is also important that the user's interaction is efficient. This is relevant to the mobile business application for two reasons

1. The user is using the application in a work context so there is a business imperative to be as efficient as possible.
2. The limited screen size and input mechanisms of handheld devices means that some interactions needs many more sequential steps than might be expected on a desktop.

Examining the complexity of the application (at design or procurement stage) is critical to ensure the application is as efficient to use as possible. To take an example from telecoms support, a human factors study of a new operator services workstation revealed the proposed design would reduce efficiency. Keystroke-level analysis, a technique for modelling the number discrete interactions that a user makes at an interface (Card et al, 1983), found inefficiencies and pointed to re-designs that were credited with saving $2 million annually (Gray et al, 1993).

St Amant et al (2007) have performed similar analyses for expert users of cell phones finding that limited input capabilities, limited screen size and lack of consistency between devices all act to undermine user efficiency. As with the aforementioned case, a detailed analysis and modelling of the discrete steps taken on a mobile device to carry out a task can reveal potential inefficiencies and design changes.

A mobile business application being used in a work context needs both *intrinsic* efficiency, in terms of how efficient it is to use the application in its own right, and *extrinsic* efficiency – it must integrate in an efficient way with the outside world, other applications on the same device, and business processes and procedures. A case study of a mobile device for engineering inspection (Dadashi et al, 2008) found that poor integration with existing work practices and excessive effort through form-filling led to severe enough usability problems that the intended user group often left the device back at the office rather than taking it out in the field as intended. Efficient integration with work tasks will be discussed further under 'Relevance to work and work context'.

Performance

The processing power of handheld devices means that most can now run without significant system response (Weiss, 2002). This is not, however, the case yet for services that require data over a network. Even as bandwidth increases, so does the amount of data being sent. System delay can have a disruptive effect on interaction – it is an irritation and at longer intervals can interrupt the flow of thought to the point where the user can become distracted. It is important to try and optimise a service to at least mask system delay.

The general rule of the thumb for system delay is that up to 0.1 seconds the user does not notice an interruption (Shneiderman, 1998). The delay will be noticeable between 0.1 and 1 second and will interrupt fluid tasks such as continuous keystrokes, or pointer movements and should be kept to a minimum. However, when users plan sequences of actions they will normally pause, and for more complex tasks a natural break will give them a chance for reflection. There will also be points where the user may be reading. At these points delays may not be an issue or even apparent to the user, so they can be used to download data in the background. Also, users tend to perceive a delay as being over as soon as they can do something useful, so it may be worth downloading partial information (e.g. text without pictures) to give the user something to work on while the rest of the data arrives. If, however, the user has to wait more than 10 seconds for the download of information, they are likely to be very frustrated – evidence from Roto and Oulasvirta (2005) suggests this might be less for mobile users with many distractions. At this point they should be given some guidance on how long the delay is going to take, and potentially have some additional feedback (e.g.tactile / vibration) to alert them the download is complete.

Accessibility

Technology is accessible when it can be used by all users, irrespective of physical, sensory or cognitive impairments. Any technology should be developed and assessed to ensure that everyone from the potential user base can interact without difficulty, and the numbers of the population that may have difficulties of using a technology may not just be those with an obvious disability. For example, there are around 2 million vision impaired and 9 million deaf and partially hearing people in the UK alone (Tiresias.org, 2008).

Designing for impaired use is particularly relevant for mobile devices because all of us will suffer a form of 'situational disability' when using mobile devices in varied contexts. For instance, trying to control a stylus on a moving train is not so dissimilar to an impairment of fine motor control, and the drowning out of sounds from a mobile device in a noisy environment is not dissimilar to a hearing impairment.

Therefore, an accessible design is a design that is good for all. The features of an accessible design include multiple input mechanisms to support users who might be limited in one or more modalities, equivalents for output (e.g. text versions of images, vibration to support audio alarms) and careful design to support high contrast and image or text discrimination. While handset manufactures make some provision for accessibility, this should also be reflected in any application that is being developed for mobile devices (either on-board or delivered as content over the network). Accessibility guidance specifically for mobile is still developing but there is guidance available from the W3C (2006) as well as the Web Content Accessibility Guidelines (WCAG, 1999).

Relevance to Work and Work Context

All of the factors discussed so far are good practice for any mobile application – business or consumer. There, are however, a number of characteristics that a business application should have, particularly if it is to be used as part of field work (as opposed to merely a mobile extension of the office).

In a detailed analysis of the use of mobile devices, Pascoe et al (2000) presented four requirements for mobile field work. They are

1. **Dynamic user configuration:** Being able to use the device in any position or conditions.
2. **Limited attention capacity:** Workers need to concentrate on their task, not on using the device, so the effort of using the mobile business application should impinge as little as possible on existing work (see Oulasvirta et al, 2005, for a more detailed discussion of the attentional demands of mobile interaction).
3. **High-speed interaction:** Field workers often cannot control the rate of work going on around them, so the application must allow them to enter data rapidly and keep up with the flow of events.

4. **Context dependency** For the field worker, issues such as time and/or location are often a critical part of the interaction. This, or other relevant context information, should be encoded automatically to cut down on the user's workload.

These findings are very similar to Dadashi et al's study of a mobile device for engineering inspection, discussed previously. Pascoe et al then go on to discuss a number of characteristics such as appropriate input, small form factor, robustness and battery life, that should be considered when developing (or procuring) a mobile application for field work.

Even when the mobile business application is supporting a more conventional office task, the work context is all important. Pew and Mavor (2007) argue that systems should not be seen in isolation, but seen as systems of systems, where one application must fit within an existing framework of technology, processes, procedures and communications. Failures of technology are often because the new system fails to integrate well with these existing elements, so a new technology, either for development or procurement, should be assessed to see how will it fits with existing work practices, both in terms of language and visual consistency, discussed previously, and with the more general work flow. The following section covers the basic process of taking a human factors approach to mobile business applications – a key phase of which is to understand the users and their work context.

A USER-CENTRED APPROACH TO MOBILE BUSINESS APPLICATIONS

While the previous section gives general guidance on the areas that should be considered when examining a mobile device for usability, there are many specific constraints the will apply to a mobile technology in actual use. It is not enough to look at the usability of any device or system in absence from the context in which it is used.

This is particularly pertinent in the case of mobile devices because a) there may be a number of contexts in which a mobile device is being used b) because the physical constraints of the device, and the very nature of its mobility, means that it is much more likely to need to support short busts of almost error-free activity (Oulasvirta et al, 2005), rather than prolonged periods of use which have more scope for trial-and-error and exploration. With that in mind, any design or deployment should be sensitive to the most critical aspects of the task the user is trying to achieve, and to ensure the design or purchased product meets user needs.

The means to achieve such a high-quality product or service is by employing a full user-centred design process. This process puts user needs and abilities, as represented by the user themselves (rather than, say, a set of marketing assumptions about what the user wants) at the heart of the development/deployment process. At first glance, this may look like another level of complexity to add to what are often already complex business processes that need to be followed when, for example, developing a new enterprise-level application. In practice, as we shall see, user-centred design practices can dovetail into existing practices such as requirements gathering, or Requests for Quotations (RFQs) in the procurement situation. With guidance they may even be achieved by the same personnel who are already involved in such activities.

There are three core phases to applying a user-centred approach

1. **Context analysis:** Where the context of use and emergent requirements are examined and understood.
2. **Specification and design:** Where the detail of requirements are refined and implemented in a way that will support the work of the end-user.

3. **Evaluation:** Where proposed designs are checked to make sure they fit with how the mobile business application will be used.

Each phase, along with the key methods involved, are described as follows. While these three phases primarily refer to design and development they also refer to procurement in two respects. First, a procurement exercise should audit that a supplier has followed such a process. Second, these techniques, particularly for context analysis and evaluation, should be part of the procurement process to ensure accurate, user-centred requirements are captured, and that the potential purchase meets those requirements.

Context Analysis

The point has been made throughout this chapter that understanding the user's context is critical to a successful product. There are a number of methods that can be used to investigate this context, and they vary in terms of their length of time they take, their naturalness and their formal outputs.

Three key methods are presented here – an ethnographic approach, task analysis and focus groups (for a comprehensive discussion of gathering user requirements see Sharp et al, 2007). They vary in their formality, and all are likely to need some input from a human factors specialist but, at the same time, they should be familiar to most people involved in understanding and structuring a customers needs, such as those in a marketing or business analysis function, and should dovetail into the elicitation phase of a traditional requirements engineering process such as Volere (Robertson and Robertson, 2006).

After the users themselves, the next most important role is those people who define at a high or detailed level, what will go into a product. Specifying high quality requirements is a proven success factor in software projects (Taylor, 2000) and the need to determine requirements early, and to make sure these requirements reflect genuine user needs, is critical to the success of any product (Cooper, 1988). The title of the role given this responsibility will vary from case to case. In some organisations this may be a product manager; in others it will be a business analyst. In a procurement or purchasing situation it may be the person specifying the Request for Quote (RFQ), Invitation to Tender (ITT) or equivalent. Whoever it is, they can integrate the methods and knowledge from human factors to ensure the requirement choices reflect real user needs and constraints.

The Ethnographic Approach

Rooted in a sociological approach to the understanding of behaviour, ethnography takes a highly observational stance towards understanding a user's context (Nardi, 1997). Typically, an observer or team of observers take a non-obtrusive position in the work context and allow users to work as normal. From often prolonged observation it is possible to build a rich picture of all of the factors that contribute to a particular work task or context. As such it is holistic, non-judgemental, centred on the point of view of the people being described and usually practiced in their natural setting.

The ethnographic approach is particularly relevant to mobile applications as its holistic approach fully appreciates all of the factors that could affect a person's work and interactions. The disadvantage is that it can be very open-ended, and its qualitative nature may need considerable interpretation before any data is ready to be integrated into a system development approach (though there are ways of targeting ethnography's use (Miller, 2000)). Overall, when the commitment to use it is made, it can be a powerful way of illuminating shortcomings in current practices and opportunities for new technology.

Task Analysis

The task analysis approach (Shepherd and Stammers, 2005) is more formal in how the output is presented. Essentially, an understanding of work or tasks is captured by a combination of observation, interview, surveys etc., and the data is structured in either a tabular format or as a Hierarchical Task Analysis (HTA) – a tree-like diagram representing a decomposition of the task flow. This may, at first glance, appear somewhat similar to standard business analysis techniques, but it is critical to note that a task analysis does not present the expected future flow of a system (it is not a diagrammatic equivalent of a use case). Instead, it can be a detailed analysis of current practice. Task analysis offers the facility to list a number of alternative methods a person will follow to meet their goals. This includes interactions with other applications or procedures that are sometimes lost in a conventional business analysis. It is particularly useful in understanding highly interactive tasks where there may be multiple steps or complex navigation required – the proposed system can be developed to integrate with the captured flow as closely as possible.

A relation of Hierarchical Task Analysis is Cognitive Task Analysis (CTA) (Crandall et al, 2006), the key difference being the emphasis on cognitive activity over observable activity. CTA comes with a number of techniques such as critical decision methods to investigate the cognitions that surround incidents that have occurred in the past, interviews and observation of key simulated tasks, and verbal protocol analysis, where the participant gives a concurrent verbal report of what they are thinking. CTA comes into its own when the task to be analysed is highly cognitive, involving a great deal of expertise, knowledge and/or problem solving. To simply observe such tasks or to make assumptions about how people are approaching them will often miss critical steps that should be considered in the design/deployment of a product. Both HTA and CTA may also point to opportunities where current work can be made more productive by the judicious introduction of the right applications.

Focus Groups

The focus group method is discussed here because it is such a pervasive marketing approach and can, if used well, be a fast and effective means of eliciting user needs and constraints (see Weilenmann (2001) for an example of its use in mobile business application development). The focus group is essentially a group interview moderated by a trained facilitator. The advantage of the focus group is that it offers a forum for people to express diverse issues, and for a consensus to build quickly. The risk is that the group will think together without dissent, and will not be as critical as might be hoped. Therefore it is possible that focus groups are best used for brainstorming rather than critique and evaluation (Jones and Marsden, 2006).

Also, for focus groups to be used effectively, participants must have a common ground through which to communicate and a clear idea of 'focus' (Hyden and Bulow, 2003). Therefore, more specific techniques can be used to guide the focus group session, such as Product Personality Profiling, Mood Boards and Visual Product Evaluation (McDonagh et al, 2002). In the software development context, it is often both economical and effective to bring users and the development team together in the form of a Joint Application Development (JAD) meeting (Wood and Silver, 1995).

Specification and Design

At this phase, high level requirements are transferred into detailed design. This may be design in terms of visual design, audio design, language and content and also as interaction flow design. While it may be tempting to leave the details of design purely to designers and developers, the reality is that they often have a partial or mistaken view of user's abilities, often over estimating their ability with technology (see Chevalier and Kicka (2006) for empirical evidence of this from Web development). Therefore, three approaches are presented here that can support user-centred development and delivery, or act as a benchmark during a procurement process – scenarios, prototyping and style guides.

Scenarios

Scenarios take information elicited at the context analysis stage, and turn it into a brief vignette or story (Carroll, 2000). These vignettes, normally written as straight-forward prose, can either describe the current state of the world, to communicate the kinds of needs and constraints that potential users may be working to, or as a future-vision story, to demonstrate the potential application. Scenarios can be powerful at the design and development phase as they give the development team a quick and easily interpreted insight into the some of the contextual conditions they should be designing for (see Nigay et al (2002) for example of using scenarios to develop a field worker mobile application). They can be converted into use cases for system requirements, and can complement the agile/extreme programming approach of developing user stories. Scenarios can be especially powerful in large-scale mobile developments or acquisitions, where middleware or even network applications need to be considered, as they serve as a common ground for all stakeholders to come together and discuss system requirements (network connectivity, security or APIs) that requires careful integration to present a seamless conceptual model to the user (Shin et al, 2006).

Prototyping

While most projects involve some form of piloting and prototyping, a stream of prototyping dedicated to human factors issues (interface design and interaction flow, in particular) can be very valuable in identifying design issues, clarifying decisions and acting as a living specification for discussion and evaluation. The main difference in a human factors approach to prototyping is

a. A specific focus on the interface, rather than this being a secondary output of a prototype to look at some underlying functionality
b. There are many forms of interface prototype, few of which need to rely on actual code (or even need a computer!) (Sharp et al, 2007).

Typically interface prototypes start at a very simplistic level, often as sketches or mock-ups in Powerpoint and graphics packages – referred to as *low-fidelity prototyping*. The advantage is that designs can be critiqued and revised very quickly, before any code is cut. An irony of prototyping, especially at the earlier stages of a project, is that sometimes rougher and simpler prototypes can leave more room for interpretation and discussion, and prove a better source of insight (Erickson, 1995).

High-fidelity prototyping occurs later in the development process, and is typically interactive. At this stage it is important to consider what platform to use for prototyping. While mobile developer environments are often readily available it is still not always easy to transfer code across onto a mobile device (particularly if it needs to be provisioned over the air e.g. as networked content). It may be appropriate to use PC-based emulation software, and this is certainly appropriate for evaluating screen flows and basic interaction properties, but if the influence of the context of use, or the capabilities of the mobile device being used, is considered critical then it is best to prototype on something that will at least mimic the appropriate platform. For more on prototyping with mobile devices refer to Jones and Marsden (2006).

Style Guides

To support consistency of look, feel, and language, a style guide should be distributed to the development team (Simpson, 1999). Depending on the resource available, and the complexity and novelty of the project, it may be appropriate to research and develop an in-house, or even project specific, style guide. At the very least, most projects require a user-centred glossary so that developers can use interface language that is both consistent, and couched in the terminology of the user. A number of companies (e.g. Nokia, Microsoft MSDN and Symbian) make their proprietary guidelines and user interface standards publicly available on the Web.

Evaluation

It is a common misconception that human factors evaluation should take place at the same time as dedicated System Acceptance Testing (SAT) or User Acceptance Testing (UAT). This is probably not helped by the frequent use of the term usability testing, to describe what should in fact be an iterative evaluation process that occurs from early prototyping through to delivery.

Much testing is *formative*. It is used to elicit information about design, and ends in a confirmatory, *summative*, phase (Pew and Mavor, 2007). Because so much evaluation occurs during development, it also differs from conventional testing in that it is not there to confirm that requirements have been met, but to examine that the manner that requirements are presented through the product/service is suitable for the intended users and their goals. It is inevitable, therefore, that the outcome of the formative evaluation will lead to changes to specification meaning that it should take place as early as possible (using the appropriate prototype), should be iterative, and that there should be scope within the project plan to handle change – this needs special consideration within an agile development approach (Chamberlain et al, 2006).

There are essentially two forms of evaluation — review and user evaluation.

Reviews

Reviews take the form of assessments against either a set of guidelines or heuristics, or a review by a human factors/HCI expert. While there is no widely accepted set of heuristics for mobile, sets of heuristics exist for software (e.g. Nielsen and Mack, 1994) and for the Web (e.g. Nielsen, 2000), that express good usability practice as a number of guidelines or qualitative benchmarks. Such a set of heuristics could form the basis of a review, and the speed an ease with which they allow usability to be assessed is appealing. The reality, however, is that results can be open to interpretation, are dependent on the skill of the reviewer, and often give a false impression (Cockton and Woolrych, 2002).

A similar approach is expert review (often they are combined so that an expert reviews against agreed heuristics), but again this can be open to interpretation and bias. In practice, reviews should only be used as an interim measure to give quick feedback on design decisions of small or medium importance. However, expert review can be appropriate, especially when the results are highly quantitative in their nature and there are clearly agreed targets – e.g. evaluating system response time over a set task, or predicting task performance by modelling factors such as number of keystrokes to complete a task (see Holleis et al, 2007, for an example of using predictive models with mobile devices).

User Evaluation

It is much more desirable to engage real users in the evaluation process. There are a number of techniques that are available but most take the form of a task-based evaluation. Here the user is given a task to achieve which they are asked to perform with the given device or application (or suitable prototype). Tasks should be developed that investigate critical tasks, known areas of human factors risk or potential issues, or that will be carried out repeatedly – bear in mind that minor but extremely frequent irritations can have just as much of a negative impact on the acceptability of an application as infrequent but catastrophic usability issues. Participants should be as representative of the intended end users as possible, and while advice about numbers of participants required varies (Barnum et al, 2003), the practical number of participants for stable results is around 5 and 10 (unless statistics are required – then the number will need to reflect the test being performed).

It is important to note that evaluation does not have to be performed with the finished article, but depending on the information required, can be performed with any form of prototype. It is usually far more cost-effective to evaluate the prototype and change the specification before coding, than evaluate the real code and have to change that.

Analysis of the results of evaluation can be qualitative or quantitative – qualitative results (normally the comments of the user, or the impressions and observations of the observer in the test session) can be time-consuming to analyse and can be open to interpretation, but has the advantage of being flexible. Often subtle, or not so subtle, additional data can be uncovered, that otherwise would be missed when only examining performance over a limited range of quantitative variables. Quantitative data can examine variables such as time to completion, error rates, or number of steps taken to completion. They are particularly useful for comparisons over time – for example, a re-design can be clearly demonstrated to be more efficient for the user because there is a tangible reduction in keystrokes. There are also a number of standardised questionnaires such as the Software Usability Scale (SUS) (Brooke, 1996), or SUMI (Kirakowski and Corbett, 1988). Questionnaires can be used at the end of an evaluation session, or distributed to a set of existing users, to quantify user impressions of system or product usability.

User evaluation tends to take place in one of two environments – in the lab or in the field. A typical usability lab allows the recording of events and comments, and is probably best suited to studies where more quantitative data is required. The disadvantage for the mobile business applications is that the lab lacks much of the realism and contextual factors that will affect the real use of the application in the workplace. In that case, a field setting should be used to evaluate the application. The arena of user evaluation is broad, and a human factors specialist should be involved to design, facilitate and analyse the results of any investigation. As a starting point, see Sharp et al (2007) or Baber (2005) for a thorough introduction to usability evaluation methods.

Take the example of developing an application for a pharmaceutical sales team to log new sales requests on a PDA style device, while out on site visits. In terms of context analysis, it would be critical to understand the typical process that the sales team goes through when working with a client. While a prolonged ethnographic study may be too in-depth for this kind of product development, it would be certainly worth seeing a few members of the sales team work in context, or at least have the opportunity to speak to them shortly afterwards, and record their task flow as a task analysis. In addition, a focus group could be used with the sales team to establish the constraints they currently face; they are likely to be time pressured and wanting to focus their attention on the client, rather than the device (similar to the issues identified by Pascoe et al (2000)). These techniques might highlight important user requirements – for example, the ability of the ordering application to pull contact details straight out of an on-device address book, or to rapidly sync up the on-device functionality with a desktop-based system back in the office. Factors such as the sales team's terminology, and the other business processes and systems the application needs to integrate with, are critical points to understand at this stage.

Next, scenarios of typical uses can be drafted. For example, a textual description of the activities a member of the sales team would currently carry out on a site visit, and the problems they face, could be distributed to business analysts and developers. It may even be worthwhile presenting scenarios in a more novel manner such as a visual storyboard. Also, early prototype designs can be drafted and shown to representatives of the sales team for discussion. Both of these measures will allow a rapid evaluation of end-user requirements, before code has been cut. Any guidance on terminology or consistency constraints can go into a style guide. As development continues and higher-fidelity prototypes, or even early versions of the real software, become available, these too should be evaluated with representatives with the sales team. Lab tests and expert review can verify that the interface is efficient to use, and meets any critical user requirements that have been identified. For example, is the structure of the on-device ordering system consistent with the navigation structure for an on-line product catalogue that the sales team uses back in the office?

As the product moves into the testing phase, summative evaluation can verify designs and particularly performance characteristics, such as system response time if the ordering system is relying on information that is being sent over a network, which can only be truly judged on a fully working system. Finally, after a product is launched (or procured) it is worth carrying out additional user evaluation, say as a focus group or observation, to make sure the sales team are accepting the tool and using it as intended, or whether they have suggestions for future iterations of the product.

FUTURE TRENDS

The future direction of human factors will be primarily shaped by new forms of mobile device and new mobile capabilities. The major emergent trend is likely to be location-aware services. While location based information is already available on mobile devices, this will be soon woven in increasingly subtle ways in to the users interaction. For example, there is significant effort directed at exploring how information from the external world can be integrated with information on mobile devices to present new levels of synergy between the digital and the physical. While this offers many new opportunities, work places and user's needs will still need to be fully understood to make sure that such emerging technology is integrated effectively (Nixon et al, 2007). Other trends will be new forms of interaction to overcome the limitations of mobile devices – such as gaze-based interaction, accelerometer (i.e. tilt-based) interaction, and new ways of integrating small mobile devices with larger devices and docks.

Following on from that, as interaction becomes increasingly mobile, so will the methods used to understand its use, especially when it comes to the evaluation of a product or service. Mobility makes laboratory-based evaluation more challenging for traditional tasks and products (Duh et al, 2005), and so approaches will need to be modified. Ethnography is likely to grow as a tool for evaluation (Crabtree et al, 2006). Models are also likely to become more advanced. Already predictive models of certain types of mobile interaction are becoming available (Parhi et al, 2006). For human factors in particular, an understanding of the physical characteristics of interaction will be important to assess not only efficiency but also potential risk of Musculo-Skeletal Disorders (MSDs) as a result of high-volume use of mobile interactive devices (see Chany et al (2008) for a comparison of the effects of muscle fatigue between conventional and mobile phone use). Safety will also be an issue for the use of mobile devices in potentially hazardous environments or safety-critical domains (Krausman and Nussbaum, 2007).

CONCLUDING COMMENTS

This chapter has given an introduction to the role of human factors in either designing a mobile product or service, or determining its suitability for procurement and implementation within an organisation. Paramount has been a rich understanding of the user, their needs and abilities, and the constraints placed upon them by their current workplace and work practices.

Table 1. Integrating human factors consideration into development (and procurement)

Traditional Waterfall	Human factors integration
Requirements Analysis	Contextual Analysis • Ethnographic approach (best suited to large projects)* • Task analysis* • Focus groups* (to compliment JAD workshops)
Design	Specification and design • Scenarios* • Low-fidelity prototyping (e.g. paper, Powerpoint) • Styleguides* (including accessibility standards, glossaries derived from contextual analysis) Summative evaluation • Early end-user feedback with low-fidelity prototypes
Code	Specification and design • High-fidelity prototyping Summative evaluation • Expert review • End user lab testing • End user field testing
Test	Formative evaluation • Expert review* • End user lab testing* • End user field testing*
Maintenance	Contextual analysis • assess on-going user acceptance to feed into training, potential re-design, and future versions*

** - these points are also applicable to procurement*

An effective understanding of this work context can be used in conjunction with the design guidance given earlier in the chapter – for example, the context analysis will help determine what language the user-base commonly uses, or what will constitute efficiency for that user group. While the ideal is to consider all factors across the whole of the development process, some factors are of particular relevance at each phase. This is outlined in Table 1, with each phase mapped to the phases of a traditional 'waterfall' development model. The waterfall model has been chosen for simplicity but the guidance is applicable to any development model. Ideally, a user centred approach should be iterative, allowing opportunity for the refinement of requirements and designs, based on analysis, prototyping and evaluation. Most important of all is taking the position that the user is a key stakeholder to be considered in the design and deployment of any business mobile application. To do so is to take a major step towards ensuring a suitable product or service that will actually deliver the business value it promises.

REFERENCES

Baber, C. (2005). Evaluation of human-computer interaction. In J. R. Wilson, & N. Corlett, (Eds.), *Evaluation of Human Work*. 3rd edn. pp. 357-388. London, UK: Taylor & Francis.

Barnum C., Bevan, N., Cockton. G., Nielsen, J., Spool, J., & Wixon, D. (2003). The "Magic Number 5": Is It Enough for Web Testing? In *Proceedings of CHI '03* (pp. 698-699). New York, NY: ACM Press.

Bergman, E., & Norman, D. A. (2000). Making technology invisible: A conversation with Don Norman. In E. Bergman (Ed.), *Information Appliances and Beyond: Information Design for Consumer Products.* San Francisco, CA: Morgan Kaufmann.

Brewster, S. A. (2002). Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland, (Eds.), *Usability Evaluation in Industry* (pp. 189-94). London, UK: Taylor & Francis.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Erlbaum.

Carroll, J. M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions.* Cambridge, MA: MIT Press.

Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a framework for integrating agile development and user-centred design. In *Proceedings of Extreme Programming and Agile Processes in Software Engineering (XP06).*

Chany, A-M., Marras, W. S., & Burr, D. L. (2008). The effect of phone design on upper extremity discomfort and muscle fatigue. *Human Factors, 49*(4), 602-618.

Chevalier, A., & Kicka, M. (2006). Web designers and Web users: Influence of the ergonomics quality of the Web site on the information search. *International Journal of Human-Computer Studies, 64*(10), 1031-1048.

Cockton, G., & Woolrych, A. (2002). Sale must end: should discount methods be cleared off HCI's shelves? *Interactions, 9*(5), 13-18.

Cooper, R. G. (1988). Predevelopment activities determine new product success. *Industrial Marketing Management, 17*, 237–247.

Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M., & Brown, B. (2006). Supporting ethnographic studies of ubiquitous computing in the wild. In *Proceedings of 2006 ACM Symposium on Designing Interactive Systems* (pp. 60-69). New York, NY: ACM Press.

Crandall, B., Klein, G., & R.Hoffman, R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA: MIT press.

Dadashi, Y., Wilson, J. R., Sharples, S., & Clarke, T. (2008). Applications of Handheld Computers for the Rail Industry. In P. Bust (Ed.), *Contemporary Ergonomics: Proceedings of the Annual Ergonomics Conference, Nottingham, 2008.* London, UK: Taylor & Francis.

Duh, H. B-L., Tan, G. C. B., & Chen, V. H. (2005). Usability evaluation for mobile device: a comparison of laboratory and field tests. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp 181-186). New York, NY: ACM Press.

Erickson, T. (1995). Notes on design practice: stories and prototypes as catalysts for communication. In J. M. Carroll (Ed.), *Scenario-based design: Envisioning work technology in system development* (pp. 37-58). New York, NY: John Wiley & Sons.

Forman, H., Lippert, S. K., & Kothandaraman, K. (2007). Understanding users' performance evaluation of IT solutions. *Industrial Marketing Management, 36*, 745-56.

Garrett, J. J. (2003). *The Elements of User Experience.* Berkeley, CA: New Riders.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Holleis, P., Kern, D., & Schmidt, A. (2007). Integrating user performance time models in the design of tangible UIs. *Proceedings of CHI'07 extended abstracts.* New York, NY: ACM Press.

Howell, M., Love, S., & Turner, M. (2006). Visualisation improves the usability of voice-operated mobile phone services. *International Journal of Human-Computer Studies, 64*(8), 754-769.

Hyden, L.C., & Bulow, P.H. (2003). Who's talking? Drawing conclusions from focus groups: some methodological considerations. *International Journal of Social Research Methodology, 5*(4) 305-321.

ISO (1998) *ISO 9241 – 11, Ergonomics Requirements for office work with visual display terminals (VDTS), Part 11: Guidance on Usability.*

Jones, M., & Marsden, G. (2006). *Mobile Interaction Design*. Chichester: John Wiley & Sons.

Kirakowski, J., & Corbett, M. (1988). Measuring User Satisfaction. In D. M. Jones, and R. Winder (Eds.), *People and Computers, IV.* Cambridge, UK: Cambridge University Press.

Krausman, A. S., & Nussbaum, M. A. (2007). Effects of wearing chemical protective clothing on text entry when using wearable input devices. *International Journal of Industrial Ergonomics, 37*(6), 525-530.

Laukkanen, T. (2007). Internet vs. Mobile Banking: comparing customer value perceptions. *Business Process Management Journal, 13*(6), 788-797.

McDonagh, D., Bruseberg, A., & Haslam, C. (2002). Visual product evaluation: exploring users' emotional relationships with products. *Applied Ergonomics, 33(1)*, 231-420.

Miller, D. R. (2000). Rapid ethnography: time deepening strategies for HCI field research. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques* (280-286). New York, NY: ACM.

Nardi, B. (1997). The use of ethnographic methods in design and evaluation. In M. Helander, T. K. Landauer, and P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed, pp. 361-366). Amsterdam, Netherlands: Elsevier.

Nielsen, J. (2000). *Designing Web Usability.* Indianapolis, IN: New Riders.

Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods.* New York, NY: John Wiley & Sons.

Nigay, L. P., Salembier, P., Marchand, T., Renevier, P., & Pasqualetti, L. (2002). Mobile and Collaborative Augmented Reality: A Scenario Based Design Approach. In *Proceedings of Mobile HCI 2002* (pp. 241-255).

Nixon, J., Sharples, S., & Jackson, M. (2007). Presenting spatial information on a mobile device: Differences in workload and performance. In *Proceedings of Human Factors and Ergonomics Society (HFES), 51st Annual Meeting.* Santa Monica, SA: Human Factors and Ergonomics Society.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York, NY: Basic Books.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. *Proceedings of CHI 2005*. New York, NY: ACM press.

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on human-computer interaction with mobile devices and services (*pp. 203-210). New York, NY: ACM press.

Pascoe, J., Ryan, N. S., & Morse, D. R. (2000). Using While Moving: HCI Issues in Fieldwork Environments. *ACM Transactions on Computer Human Interaction*, *7*(3). 417-437.

Payne, S. J. (2003). Users' mental models: The Very Ideas. In J. M. Carroll (Ed.), *HCI Models, Theories and Frameworks: Toward a multidisciplinary science*. San Francisco, CA: Morgan Kaufman.

Pew, R. W., & Mavor, A. S. (2007). *Human-System Integration in the System*

Development Process: A new look. Washington, DC: National Academies Press.

Robertson, S., & Robertson, J. (2006). *Mastering the Requirements Process,* 2nd edn. Boston, MA.: Addison-Wesley.

Roto, V., & Oulasvirta, A. (2005). Need for Non-Visual Feedback with Long Response Times in Mobile HCI. *Proceedings of the 14th International World Wide Web Conference*. New York, NY: ACM.

Sharp, H., Rogers, Y., & Preece, J. (2007). *Interaction Design: Beyond Human-Computer Interaction 2nd Ed*. London, UK: John Wiley & Sons.

Shepherd, A., & Stammers, R. B. (2005). Task analysis. In J. R. Wilson, & N. Corlett, (Eds.). *Evaluation of Human Work*. 3rd edn. London, UK: Taylor & Francis.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer interaction,* 3rd edn. Reading, MA: Addison Wesley Longman.

Shin, D., Kim, J., Ryu, S., Oh, D., Lee, J., & Kang, M. (2006). Scenario Decomposition Based Analysis of Next Generation Mobile Services. *Proceedings of the International Conference on Computational Science, (2)* 977-984.

Simpson, N. (1999). Managing the use of style guides in an organisational setting: Practical lessons in ensuring UI consistency. *Interacting with Computers, 11*(3), 323-351.

St. Amant, R., Horton, T. E., & Ritter, F. E. (2007). Model-based evaluation of expert cell phone menu interaction. *ACM Transactions on Computer-Human Interaction, 14*(1).

Taylor, A. (2000). IT projects: sink or swim. *The Computer Bulletin,* Jan, 24-26.

Tiresias.org (2008) Retrieved 28th Feb 2008, from http://www.tiresias.org/guidelines/.

Väänänen-Vainio-Mattila, K., & Ruuska, S. (2000). Designing Mobile Phones and Communicators for Consumers' Needs at Nokia. In E. Bergman (Ed.), *Information Appliances and Beyond: Information Design for Consumer Products.* San Francisco, CA: Morgan Kaufmann.

W3C (2006). *Mobile Web Best Practices 1.0.* Retrieved 28th Feb 2008, from http://www.w3.org/TR/mobile-bp/

WCAG (1999). *Web Content Accessibility Guidelines 1.0.* Retrieved 28th Feb 2008, from http://www.w3.org/TR/1999/WAI-WebCONTENT-19990505/

Weiss, S. (2002). *Handheld Usability.* Chichester, UK: John Wiley & Sons.

Weilenmann, A. (2001). Negociating use: Making sense of mobile technology. *Personal and Ubiquitous Computing, 5*(2), 137-145.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). *An Introduction to Human Factors Engineering,* 2nd edn.. Upper Saddle River, NJ: Pearson Education.

Wilson, J. R. (2000). Fundamentals of Ergonomics in theory and practice. *Applied Ergonomics, 31*(6), 557-567.

Wilson, J. R., & Corlett, N. (Eds.), (2005). *Evaluation of Human Work,* 3rd edn. London, UK: Taylor & Francis.

Wood, J., & Silver, D. (1995) *Joint Application Development*, 2nd edn. New York, NY: John Wiley & Sons.

KEY TERMS

Accessibility: Designing technology so that is suitable for all intended users, regardless of any physical, sensory or cognitive impairment.

Context Analysis: A critical stage in a user-centred approach that investigates and analyses the work context of potential users, in order to understand the factors that will influence the usability of a product or service.

Evaluation: An ongoing activity executed throughout the development or procurement process, where a mobile business application is evaluated, either as a prototype or as a finished product. Ideally this is conducted with typical end-users, to ensure it integrates in a usable manner with their existing working practices.

Human Factors: A domain offering research and guidance on how to design work, and the tools of work, to provide the best match with work goals, people's needs, and human physical and cognitive capabilities.

Prototyping: An activity of designing early mock-ups of the user interface – either as low-fidelity or high-fidelity prototypes – to allow discussion and reflection on design choices, and for use in evaluation.

Usability: The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

User-Centred Design: An approach to designing and developing technology that places the user, their needs, capabilities and limitations at the heart of the development process.

Section V
Telecommunications Technology and Applications

Chapter XLIII
Towards QoS-Inferred Internet

Mohamed Boucadair
France Telecom-Orange Labs, France

Pierre Levis
France Telecom-Orange Labs, France

Pierrick Morand
France Telecom-Orange Labs, France

ABSTRACT

IP networks are the federative transport networks for a large set of emerging services. These services demand hard guarantees in term of the service availability, experienced Quality of Service (QoS) and robustness. Moreover, to be able to reach customers on a large scale, most of these services should be deployed with an Inter-domain scope. In order to meet QoS requirements of these services in an inter-domain context, several issues should be solved. This chapter focuses on two issues: provider-to-provider agreements and enhancements to inter-domain routing protocol to convey QoS-related information. A concept called Meta-QoS-Class is introduced together with an enriched version of Border Gateway Protocol. This chapter provides a framework suitable for the promotion of QoS-enabled services with an inter-domain scope: the Parallel Internet. This concept is a viable way for the management of IP resources so as to deliver end-to-end QoS-enabled services.

1 INTRODUCTION

1.1 Background and Focus

Quality of Service (QoS) is one of the topics that succeeded to federate various Internet actors. Indeed, experts from both academia and industry have contributed to promote QoS-centric solutions within standardisation bodies, such as IntServ (Integrated Services) or DiffServ (Differentiated Services) Initiatives within the IETF (Internet Engineering Task Force)(IETF is the main IP standardisation body. This is the place where IP-related issues, protocols and architectures are designed.). Most of that effort has been put into solving intra-domain (i.e. within the IP infrastructure owned and managed by the same IP Network Provider) specific issues. Moreover, some of the QoS proposals have been adopted by Network Providers and even have been activated in their operational networks. Before it is conceivable to have an operational deployment of end-to-end QoS services (i.e. across domains owned by plethora of IP Network Providers), a number of issues still appear to need further elaboration as identified in Huston (2000). In fact, there is a strong lack of inter-domain QoS investigation and few "big picture" solutions have been proposed. This chapter's ambition is to contribute to this area and advocate for the introduction of solutions dealing with inter-domain QoS delivery services.

For readers who are not familiar with QoS notion, we provide the ITU definition of Quality of Service as defined in E.800. Indeed, QoS is defined as "the collective effect of service performance which determines the degree of satisfaction of a user of the service".

As stated in Deleuze (1996), concepts should always be created in relation to specific problems. This section focuses exclusively on inter-domain QoS service delivery issues since intra-domain ones have benefited from a large amount of effort in both academia and standardisation bodies. This paper proposes novel means and mechanisms to ease extending intra-domain QoS capabilities beyond the boundaries of a single Network Provider. A concept denoted by Meta-QoS-Class and an extended BGP (Border Gateway Protocol) protocol to be able to convey QoS-related information (see Boucadair (2005)) are introduced. A framework is also described to promote the notion of Parallel Internet (refer to Boucadair (2006)). An implementation example based on joint exploitation of QoS-Enhanced BGP and Meta-QoS-Class is provided.

The emergence of Parallel Internet is essentially driven by the needs of end-to-end applications mainly in terms of QoS and resilience requirements. Owing to the adoption of Meta-QoS-Class concept, a coherent and consistent QoS treatment will be experienced by IP flows when crossing several Autonomous Systems managed by distinct Network Providers. QoS-Enhanced BGP ensures a global reachability within a Meta-QoS-Class routing plane, since several routing tables will be maintained, each per Meta-QoS-Class. Parallel Internet is a step forward to implement end-to-end service differentiation and facilitate the emergence of new business models suitable for the services of the future. Adhering to this approach, telecom players will clarify business roles and frontiers will be abolished between Service Providers.

Within this chapter, we introduce a QoS-Inferred Parallel Internet. Availability and robustness issues are out of scope of this paper.

1.2 Structure

This paper is structured as follows. Section 2 analyzes provider-to-provider QoS agreements suitable for a global QoS-enabled Internet and introduces the concept of Meta-QoS-Class. This concept drives and federates the way QoS inter-domain relationships are built between Network Providers. Section 3 is dedicated to QoS-Enhanced BGP proposal, an enriched version of BGP. This section describes attributes to convey QoS-related information in inter-domain routing protocol and presents also novel logics to compare and select QoS-Inferred BGP routes. Finally, Section 4 advocates for the introduction of Parallel Internet so as to offer a global infrastructure for delivering end-to-end QoS services.

2 CONSIDERATIONS OF PROVIDER-TO-PROVIDER AGREEMENTS FOR INTERNET-SCALE QOS

This section shows that for the sake of scalability, providers need not be concerned what occurs more than one hop away (from their Autonomous System) when they negotiate inter-domain QoS agreements. They should base their agreements on nothing but their local QoS capabilities and those of their direct neighbors. This analysis leads us to define terminology relevant to inter-domain QoS models. We also introduce a concept denoted by Meta-QoS-Class (MQC).

2.1 Assumptions and Requirements

We assume that provider-to-provider QoS agreements are negotiated only for two adjacent domains that are directly accessible to each other (immediate neighbors). We also assume that these neighbors are BGP peers (more details about BGP are documented in Rekhter (2006)). This pair-wise peering is logical. We exclusively envisage here QoS solutions that are suitable for the global Internet. This leads us to define some assumptions and requirements:

- A solution should have the ability to establish QoS communications between any two end-users, in the way we are used to with Internet best-effort communications. That means we do not target QoS introduction in a limited set of domains as for VPNs (Virtual Private Network), but we envisage true global Internet communications.
- A solution should be usable as soon as it begins to be deployed. We should not have to wait for a complete (Internet large) deployment, but rather we should be able to actually set up QoS guaranteed solutions even when the solution is only deployed in a small set of domains.
- A solution should be scalable in order to allow a global deployment to almost all Internet domains.
- If there is no path available within the requested QoS to reach a destination, this destination must remain reachable through the best-effort service.
- The best-effort service should remain the pre-eminent service as a consequence of the end-to-end argument as presented in Saltzer (1984).

No specific requirements are placed on the intra-domain traffic engineering policies and the way they are enforced. A Service Provider may deploy any technique to ensure QoS inside its own network. We only assume that QoS capabilities inside a Service Provider's network can be represented as local-QoS-Classes (l-QCs). When crossing a domain, traffic experiences conditions characterized by the values of delay, jitter, and packet loss rate that correspond to the l-QC selected for that traffic within that domain. Capabilities can differ from one Service Provider to another by the number of deployed l-QCs, by their respective QoS characteristics, and also by the way they have been implemented and engineered.

2.2 Problems with Provider-to-Provider QoS Agreements based on SP Chains

We analyze provider-to-provider QoS agreements based on guarantees that span several domains and emphasis their vulnerabilities. In this case, the basic service element that a Service Provider offers to its neighboring Providers is based on the notion of SP chains. We point out several weaknesses of such an approach, especially the SP chain trap problem that leads to the so-called Internet glaciation era.

2.2.1 SP Chains

In Figure 1, **SPn** offers a QoS service to **SPn-1** for a destination located in **SP1**. Provider **SPn** guarantees Provider **SPn+1** the level of QoS for crossing the whole chain of Providers' domains **(SPn, SPn-1, SPn-2, ...,SP1)**. **SPi** denotes a Service Provider as well as its domain. The top of the

Figure 1. SP chains

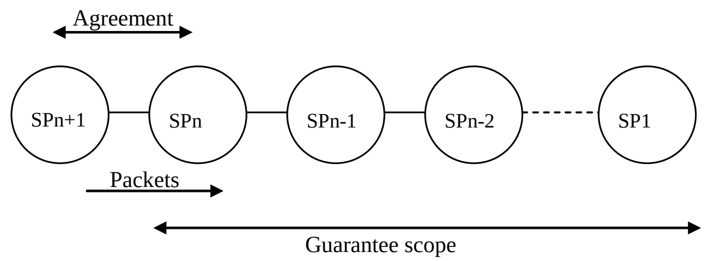


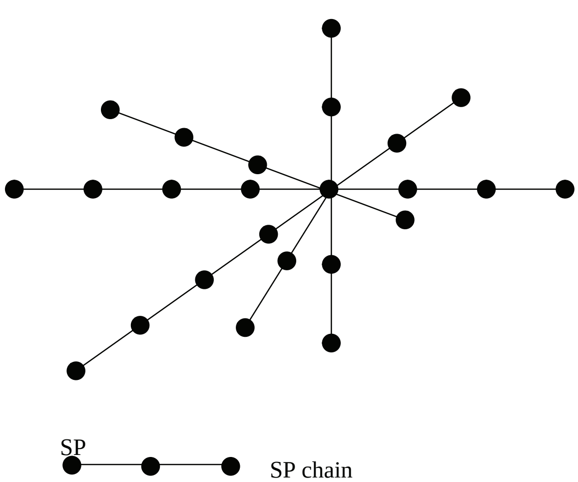
figure is the SP-oriented view, the ordered set of Service Providers **(SPn, SPn-1, SPn-2, ...,SP1)** is called an SP chain. The bottom of the figure is the domain-oriented view.

2.2.2 SP Chain Trap Leading to Glaciation

In Figure 1, **SPn** implicitly guarantees **SPn+1** the level of QoS for the crossing of distant domains like **SPn-2**. SP chains are bound to proliferate. A Service Provider is, in this context, likely to be part of numerous SP chains. It will see the level of QoS it provides guaranteed by many SPs it has maybe even never heard of.

Any change in a given agreement is likely to have an impact on numerous external agreements that make use of it. A Service Provider sees the degree of freedom to renegotiate, or terminate, one of its own agreements being restricted by the large number of external (to its domain) agreements that depend on it. This is what is referred to as the "SP chain trap" issue. In Figure 2 the central SP is trapped in several SP chains. This way of setting provider-to-provider agreements is not appropriate for worldwide QoS coverage as it would lead to glaciation phenomena, causing a completely petrified QoS infrastructure, where nobody could renegotiate any agreement.

Figure 2. SP chain traps



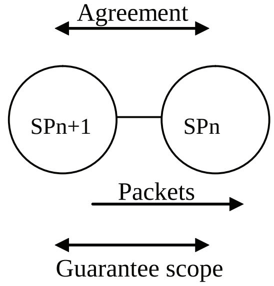
If a QoS-enabled Internet is deemed desirable, with QoS services available potentially to and from any destination, any solution must resolve the aforementioned weaknesses and scalability problems, and find alternate schemes for provider-to-provider agreements.

2.2.3 Single Domain Covering

Due to the vulnerabilities of the SP chain approach, we assume provider-to-provider QoS agreements should be based on guarantees covering a single domain. A Provider guarantees its neighbors only the crossing performance of its own domain. In Figure 3, **SPn** guarantees the Provider **SPn+1** only the QoS performance of the **SPn** domain. This approach, bringing clarity and simplicity into inter-domain relationships, is better suited for a global QoS Internet than that based on SP chains.

It is very important to note that the proposition to limit guarantees to only one domain hop applies exclusively to provider-to-provider agree-

Figure 3 Provider-to-Provider QoS agreement



ments. It does not in any way preclude end-to-end guarantees for communications. The simple fact that SP chains do not exist makes the AS chain trap problem and the associated glaciation threat vanish. The liability issue is restricted to a one hop distance. A Network Provider is responsible for its own domain only, and is controlled only by all the neighbors with whom it has a direct contract.

2.2.4 Binding l-QCs

When a Network Provider wants to contract with another Network Provider, the main concern is to decide which l-QC(s) in its own domain it will bind to which l-QC(s) in the neighboring downstream domain. The l-QC Binding process becomes the basic inter-domain process.

If one l-QC were to be bound to two (or more) l-QCs it would be very difficult to know which l-QC the packets should select. This could imply a flow classification at the border of the domains based on granularity as fine as the application

Table 1. l-QC Binding

	Upstream domain		Downstream domain
Allowed	l-QC21	→	l-QC12
Allowed	l-QC23 l-QC24	→	l-QC13
Forbidden	l-QC21	→	l-QC13 l-QC14

flows. For the sake of scalability we assume one l-QC should not be bound to several l-QCs (Levis, 2004). On the contrary, several l-QCs can be bound to the same l-QC, in the way that l-QC23 and **l-QC24** are bound to **l-QC13** in Table 1.

A provider decides the best match between l-QCs based exclusively on:

- *What it knows about its own l-QCs;*
- *What it knows about its neighboring l-QCs.*

It does not use any information related to what is happening more than one domain away.

Despite this one hop, short-sighted approach, the consistency and the coherency of the QoS treatment must be ensured on an l-QC thread formed by neighboring bound l-QCs. Packets leaving a domain that applies a given l-QC should experience similar treatment when crossing external domains up to their final destination. A Network Provider should bind its l-QC with the neighboring l-QC that has the closest performance.

The criteria for l-QC binding should be stable along any l-QC thread. For example, two SPs should not bind two l-QCs for minimizing the delay whereas further on, on the same thread, two other SPs have bound two l-QCs for minimizing errors. Constraints should be put on l-QC QoS performance parameters to confine their values to an acceptable and expected level on an l-QC thread scale. These constraints should depend on domain size, for example restrictions on delay should authorize a bigger value for a national domain than for a regional one. Some rules must therefore be defined to establish in which conditions two l-QCs can be bound together. These rules are provided by the notion of Meta-QoS-Class (MQC).

2.2.5 Meta-QoS-Class concept

A Meta-QoS-Class provides the limits of the QoS parameters two l-QCs must respect in order to be bound together. Meta-QoS-Class QoS parameters will typically give qualitative or quantitative (boundaries or intervals) information for the **(D, J, L)** parameters, where **D** is the one-way transit delay defined in Almes (1999), **J** is the one-way transit delay variation or jitter described in Demichelis (2002), and **L** is the packet loss rate introduced in Almes (1999). A Meta-QoS-Class could be focused on a single parameter (e.g. suitable to convey low delay sensible traffic). Several levels could also be specified depending on the size of the Network Provider, for instance a small domain (e.g. regional) needs lower delay than a large domain (e.g. national) to match a given Meta-QoS-Class.

Two l-QCs can be bound together if, and only if, they conform to the same Meta-QoS-Class.

A Network Provider goes through several steps to extend its internal l-QCs through the binding process. Firstly, it classifies its own l-QCs based on Meta-QoS-Classes. A Meta-QoS-Class is used as a label that certifies the support of a set of applications that bear similar network QoS requirements. It is a means to make sure that an l-QC has the appropriate QoS characteristics to convey the traffic of this set of applications. Secondly, it learns about available Meta-QoS-Classes advertised by its neighbors. To advertise a Meta-QoS-Class, a Network Provider must have at least one compliant l-QC and should be ready to reach agreements to let neighbor traffic benefit from it. Thirdly, it contracts an agreement with its neighbor to send some traffic that will be handled according to the agreed Meta-QoS-Classes.

3 QOS-ENHANCED BGP

The deployment of Internet is a success owing to a fruitful cooperation between several actors mainly Service Providers, standardisation bodies, regulators and equipment manufacturers. Network Providers have deployed standard inter-domain routing protocols in order to convey reachability information between their domains. The existence of such standard protocols has facilitated interconnection between distinct Autonomous Systems (AS) and then allowed to reach remote destinations located beyond the boundaries of a single Network Provider's domain. Nowadays, the required information to be exchanged between Network Providers, i.e. between their respective domains, is different from what could be exchanged owing to existing inter-domain routing protocols. Indeed, reachability information should be richer than what is exchanged via current routing protocols and should provide routers with pertinent information in order to drive the route selection decision-making process. Such information could be for instance the QoS that will be experienced along a given path within the context of QoS delivery services which are seen as a part of future Internet services Atkinson (2004). From this perspective, it is obvious that Network Providers have to evolve and update the protocols they are used to deploy in their domains in order to meet the new requirements and then to be able to offer new added value services.

This chapter introduces a proposal allowing exchanging QoS-related information between adjacent ASes (detailed specifications are documented in Boucadair (2005)). This section describes a proposal that benefits from the extensibility capability offered by the Border Gateway Protocol (BGP) and that meets a set of generic requirements described as follows.

3.1 Taxonomy

The QoS-related information exchange occurs either at the service level or at the routing level. The place this exchange occurs depends on the deployed inter-domain QoS delivery solution. Two groups of QoS delivery solutions have been identified and are detailed hereafter:

- **Group 1:** The first group of solutions requires propagating only an identifier that has been agreed during the interconnection agreement negotiation phase (see Figure 3). Additional QoS performance characteristics were negotiated but not exchanged in the routing level.
- **Group 2:** The second group requires the propagation of a set of QoS performance characteristics associated with an identifier. The nature of the QoS-related information to be exchanged has to be agreed during the interconnection agreement negotiation phase.

3.2 Requirements

As stated previously, BGP is the inter domain routing protocol used to interconnect domains. This protocol is widely deployed and activated in a big range of network nodes. One of the risks to be taken into account when introducing new services such as inter-domain QoS ones is to preserve backward compatibility. Therefore, in order to ease introduction of these added value services, it is recommended to reuse existing protocols and systems. Nevertheless, these protocols should be evolved and enhanced with additional features in order to achieve these new service objectives. As far as inter domain routing is concerned, we will reuse BGP and define new features in order to convey QoS-related information between adjacent domains. This information could only be reduced to a DS (Differentiated Services, Blake (1998)) code point or be a set of QoS performance characteristics. In the remaining part of this paper, we will refer to this enhanced BGP as q-BGP (QoS-Enhanced BGP) protocol.

When designing extensions to be added to classical BGP, the following requirements are to be taken into account:

- The q-BGP protocol should, as far as possible, be able to operate independently of the inter-domain QoS delivery service it serves.
- q-BGP should be able to propagate topology changes without any significant impact on the existing best-effort based network infrastructure.
- q-BGP should be able to support all kind of services based on an exchange of QoS-related information especially to serve the two groups described previously.

3.3 Provisioning QoS-Enhanced BGP

QoS-Enhanced BGP encloses in its messages QoS performance characteristics that could be advantageously taken into account by the q-BGP route selection process to select an optimal path. This would enable to tune the route selection process in order to select routes according to more sophisticated routing policies (e.g. route with highest available rate and lower delay). The QoS information inserted in QoS-Enhanced BGP messages could be of different nature. It could be (1) administratively enforced. In that case it would not change too frequently. Or, it could (2) be much more dynamic (result of an active measurement for instance). In that case the frequency of changes could be much higher.

Administrative setting of QoS values could be achieved either statically or periodically. If these values are set statically, the behavior of QoS-Enhanced BGP will be static and the route selection process will choose the same route. The QoS-related information does not bring major added-value to the final behavior of the route decision-making process and freezes the state of the inter-domain routing. Nevertheless, in the case of QoS performance characteristics values are set periodically or dynamically, Network Providers will deploy mechanisms that monitor the network and then guide the setting of these values. QoS-Enhanced BGP will be provided by accurate information in order to select the optimal path. The frequency between two QoS-Enhanced BGP router configuration operations in an administrative scheme should not be too small and could be very small in the dynamic scheme. In case of dynamic setting scheme, the risk is to impact routing table stability and probably introduce oscillation phenomena.

3.4 QoS-Enhanced BGP Specifications

QoS-Enhanced BGP has several behaviors depending on the nature of the QoS-related information carried by its messages. If q-BGP messages carry only a QoS Class (QC) identifier (this identifier could be a ToS code-point or a proprietary identifier), offline traffic engineering functions are certainly complex but the q-BGP route selection process complexity is reduced. This complexity increases when a set of QoS characteristics are associated with each QC identifier. The route selection process can use either the QC-identifier for all services that take part of "group-1" or the QC-identifier and QoS performance characteristics for solutions belonging to "group-2". This section introduces q-BGP attributes, their processing and describes examples of route selection processes.

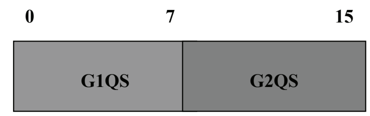
3.2.1 QoS attributes

This section introduces two attributes "**QoS Service Capability**" and **QoS_NLRI**.

3.4.1.1 QoS Service Capability

It is useful for a q-BGP peer to know the capabilities of a q-BGP neighbor with respect to the q-BGP protocol extensions (Chandra, 2002; Bates, 2007). Capabilities exchange is achieved owing to the specification of a new optional parameter. This parameter is included as an optional parameter of the **OPEN** message. In order to indicate that a given inter-domain QoS delivery solution belongs to a given group (either "group-1" or "group-2"), we introduce a new capability called "**QoS Service Capabilities**". A q-BGP speaker should use this capability advertisement in order to indicate the group to which an offered inter-domain QoS delivery solution belongs to, so that its peers can deduce if they can use the "QoS service"-related attributes with this q-BGP peer. The capability value field is encoded as shown in Figure 4.

Figure 4. QoS service capability attribute



1. The first octet is set to **0xFF** if an offered inter-domain QoS delivery solution that belongs to "group-1" is supported;
2. The second octet is set to **0xFF** if an offered inter-domain QoS delivery solution that belongs to "group-2" is supported.

3.4.1.2 QoS_NLRI attribute

As described previously, both "group-1" and "group-2" solutions need to exchange a QC identifier. This identifier is used to differentiate between the extended QCs that have been bought to service peers. In addition to this identifier, solutions belonging to "group-2" need to exchange also a set of QoS performance characteristics. To achieve this goal, a new attribute called **QoS_NLRI** is introduced. For "group-1", the format of this **QoS_NLRI** attribute is illustrated in Figure 5 and the one for "group-2" is shown in see Figure 6. For more information about this attribute, readers are invited to refer to Boucadair (2005)

Figure 5. QoS_NLRI attribute for group-1

0 1 2 3 4 5 6 7
QoS class identifier (1 octet)
Address Family Identifier (2 octets)
Subsequent Address Family Identifier (1 octet)
Next Hop Network Address length (1 octet)
Network Address of Next Hop (variable)
NLRI information (variable)

Figure 6. QoS_NLRI attribute for group-2

0 1 2 3	4 5 6 7
QoS Information length (1 octet)	
QoS Information Code	QoS Information Sub-code
QoS Information Value (2 octets)	
QoS Information Code	QoS Information Sub-code
QoS Information Value (2 octets)	
:	
QoS Information Code	QoS Information Sub-code
QoS Information Value (2 octets)	
QoS class identifier (1 octet)	
Address Family Identifier (2 octets)	
Subsequent Address Family Identifier (1 octet)	
Length of Network Address of Next Hop (1 octets)	
Network Address of Next Hop (variable)	
NLRI information (variable)	

3.4.2 Processing q-BGP attributes

As described in section 3.4.1.1, q-BGP peers exchange their respective capabilities through capability negotiation procedure. As a consequence, q-BGP peers will conclude if they both support **QoS_NLRI** attribute or not. If a q-BGP speaker does not support capability negotiation feature, it will be hard to know in advance its behaviour when receiving **QoS_NLRI** attribute. Therefore, three scenarios should be examined in order to describe the processing of **QoS_NLRI** attribute by a given q-BGP speaker:

- **Scenario 1:** If a q-BGP speaker does not support BGP capability feature, no **QoS_NLRI** should be sent to this peer.

- **Scenario 2:** If a q-BGP speaker does not support "**QoS Service Capability**", no **QoS_NLRI** should be enclosed in BGP messages destined to this peer.
- **Scenario 3:** Both q-BGP peers support "**QoS Service Capability**". In this case, q-BGP peers use **QoS_NLRI** attribute. The variant of **QoS_NLRI** attribute that will be used depends on the nature of the deployed inter-domain QoS delivery solution, either it is a "group-1" or "group-2".

3. When sending a **QoS_NLRI** attribute, the local q-BGP speaker should set the QC identifier field to the identifier of extended QC on the corresponding inter-domain link. In addition, if it is a "group-2" solution and if the q-BGP peer supports "group-2" QoS delivery solution, the local q-BGP speaker should set the value of "**QoS Information value**" field(s).
4. When receiving a **QoS_NLRI** attribute, q-BGP speaker applies its inbound policies to grant the received announcement depending on QC binding list. The local q-BGP speaker gets the value of the "**QoS Class Identifier**" enclosed in the **QoS_NLRI** of the received announcement and checks if there is a binding entry.
5. If there is no entry in the binding list: the local q-BGP speaker drops the received announcement.
6. If there is an entry in the binding list: the local q-BGP speaker updates the values of "**QoS Information value**" enclosed in the **QoS_NLRI** with the local QC ones.

3.4.3 QoS-Enhanced BGP Route Selection Process

As far as QoS-related information is conveyed in BGP **UPDATE** messages, the route selection process should take into account this information in order to make a choice and make a tie-break between equal paths and determine the one(s) to be stored in the local RIB (Routing Information Base). This process differs between solutions that belong to "group-1" or "group-2" as described in the following two sub-sections.

3.4.3.1 Group-1 Route Selection Process

For inter-domain QoS-delivery solutions that belong to "group-1" only the identifier of the extended QC is to be taken into account in order to choose a path that will be stored in the local RIB. The unique modification to be added to the classical route selection process is to identify routes that serve the same destination with similar extended QCs. Local policies could be configured by each Network Provider in their ASBRs (Autonomous System Border Routers). From this perspective, the pseudo code of the BGP modified route selection process is as follows:

```
Identify the received routes that serve the
same destination;
Consider the routes with similar extended
QCs;
Apply local policies (e.g. prefer a given
origin AS, cost,…).
If only one route has been returned Store
this route in the RIB;
If more than one route has been returned
Apply the classical BGP route selection
process.
```

3.4.3.2 Group-2 Route Selection Process

For inter-domain QoS-delivery solutions that belong to "group-2", q-BGP **UPDATE** messages carry QoS performance characteristics together with a QC identifier. QoS-Enhanced BGP route selection process exploits enclosed QoS performance characteristics in order to determine the path that will be stored in the local RIB. The following is the pseudo code of the new route selection process:

```
Identify routes that serve the same des-
tination
Consider routes that have the same QoS
class identifier
Compare the QoS performance characteris-
tics associated with resulting routes with
```

```
respect to a given comparison logic
Return the route that optimizes the QoS
performance characteristic
if more than one route has been returned,
apply the classical BGP route selection
process
```

3.4.3.2.1 QoS Comparison Logic

In this section, we discuss several approaches for comparing sets of QoS values enclosed in q-BGP messages. Consider two QoS tuples **X** and **Y**. These tuples consist of both the attributes (e.g. delay, jitter, loss rate) and their values. Let the tuples consist of QoS attributes **A**, **B** and **C**, and let the QoS tuple **X** have the values **(Ax, Bx, Cx)** and let QoS tuple **Y** have the values **(Ay, By, Cy)**.

Then to compare the two QoS tuples **X** and **Y**, a number of mechanisms can be adopted. To generalize the discussion, here we assume that "**P>Q**" means that **P** is better than **Q**, irrespective of whether we are comparing bandwidth values (where a higher numerical value represents a better level of QoS) or delay values (where a lower numerical value represents a better level of QoS). The proposed methods are as follows:

- **Lexicographical ordering method:** The QoS attributes are compared in strict order. Thus if **Ax > Ay** then **X** is better than **Y**, irrespective of the relative values of **Bx**, **By**, **Cx** or **Cy**. If **Ax = Ay** then the second QoS attributes are compared: if **Bx > By** then **X** is said to be better than **Y**. We refer to the route selection process that uses this QoS comparison logic as priority-based route selection process.
- **Simultaneous comparison method:** **X** is better than **Y** if **Ax > Ay** and **Bx > By** and **Cx > Cy**. Similarly, **Y** is better than **X** if **Ay > Ax** and **By > Bx** and **Cy > Cx**. This approach does not define a result if some of the QoS attributes **A**, **B**, **C** of one tuple are better than the second tuple but some of the QoS attributes are worse. For example, if **Ax > Ay** while **By > Bx** then the result of the comparison of **X** with **Y** is undefined. This approach is not recommended to be used as QoS comparison logic in route selection process implementations.
- **Weighted ordering method:** The QoS attributes are normalized to create dimensionless values, and summed. This results in a single value for each QoS tuple, which can be compared to determine which tuple is better. The dimensionless values could additionally be weighted so as to prefer one attribute over others. The advantage of this approach is that it potentially allows a wider range of QoS metrics to be fairly compared, for example "a low delay route with reasonable bandwidth".

3.4.3.2.2 Priority-Based Route Selection Process

When a route selection process implements lexicographical comparison logic, priority values must be assigned to QoS performance characteristics. Then, the comparison of available routes should be based on the use of the priority value that has been affected to each QoS performance characteristic. The priority ordering of the QoS performance characteristics could be commonly understood by Network Providers or only configured by each Network Provider. Therefore, the pseudo code of the priority-based route selection process algorithm is as follows:

```
Identify routes that serve the same destination
Consider routes that have the same QoS class identifier
Consider the QoS performance characteristic that has the highest priority, and return the routes that optimise that QoS performance characteristic
   i. If only one route is returned store this route in the local   RIB
    ii. If more than one route are returned:
      a. Exclude the QoS performance characteristic that has been used in the step 3 from the list of QoS performance characteristics.
```

```
        b. If there are no remaining QoS
performance characteristics, go to step 4
       c. Else, go to step 3
If more than one route has been returned,
apply the classical BGP route selection
process.
```

To illustrate this route selection process, we provide the flowing example: Suppose that a q-BGP listener has received the following routes for reaching the same destination **D1**. Each of these routes is associated with a set of QoS performance values as follows:

- **R1**: **minimum-one-way-delay = 150ms**, **loss-rate=9%**
- **R2**: **minimum-one-way-delay = 90ms**, **loss-rate=4%**
- **R3**: **minimum-one-way-delay = 100ms**, **loss-rate=5%**
- **R4**: **minimum-one-way-delay = 200ms**, **loss-rate=1%**

If the q-BGP router is configured to prioritize **minimum-one-way-delay**, the selected route is **R2**. But if the q-BGP router is configured to prioritize **loss-rate**, the selected route is **R4**.

3.4.3.3 Processing of route with "QoS holes"

Lets suppose now that a q-BGP router has received from its peers, the following routes for reaching the same destination **D1**. The received routes enclose the following QoS performance values as detailed:

- Route R1: **QoS1=90ms**, **QoS3=150ms**, **QoS4=5%**
- Route R2: **QoS2=30ms**, **QoS3=153ms**, **QoS4=1%**
- Route R3: **QoS1=120ms**, **QoS2=100ms**, **QoS3=60ms**, **QoS4=3%**
- Route R4: **QoS2=90ms**, **QoS3=50ms**, **QoS4=8%**

The aforementioned routes enclose different QoS performance characteristics. The issue is how to compare these routes?

This problem could be caused by a Network Provider which does not support some QoS parameters or could be due to a mis-configuration. The risk, in both cases, is that Network Providers will not advertise QoS-related information since if only one domain in the chain does not implement a QoS parameter it will introduce a "QoS hole" in all routes that it will advertise and then impact the announcement of this route for downstream domains. In order to solve this issue, the hereafter alternatives could be implemented:

- Add a new control level in the definition of l-QC. This consists in defining "**Mandatory**" and "**Optional**" attributes. A "**Mandatory**" QoS information is a parameter that must be present in the **QoS_NLRI**. In the case it is missing the announcement will be dropped by a q-BGP receiver. The announcement is not dropped if an "**Optional**" QoS Information is missing. Nevertheless, the problem of ensuring route selection consistency when "**Optional**" parameters are missing is unsolved. For solving this case, one of options details under "Solution 2" bullet could be implemented. It is clear that all Network Providers should have the same understanding of the "**Mandatory**" and the "**Optional**" parameters in order to preserve a consistent and coherent treatment in crossed domains. We refer to this as "Option 1" in Figure 7.
- No additional control level is introduced in the local QoS Class definition. In this second solution, the risk is that Network Providers won't advertise routes with required QoS information that should be used as guidance to meet service needs. As a consequence, "group-2" solutions become as "group-1" ones because there is no control regarding

the enclosed QoS information. When a QoS parameter is missing, two options could be considered.

- **Option 2:** Discard unvalued routes but keep them all if they are all unvalued. In this case, the priority criterion is respected and the comparison between routes is consistent. But, the risk is that some destinations could be unreachable if received routes do not enclose higher priority QoS performance characteristics.
- **Option 3:** Always keep routes with unvalued parameter, but perform selection for the remaining routes. The route selection process compares between routes that have valued the QoS parameter used as criterion in a given step. If there still have equal routes, all routes are considered and the route selection process checks for the route that encloses the next QoS parameter to be used as criterion of its selection even if these routes were not present in the previous step. The inconvenient of this option is that the priority criterion is not satisfied.

To illustrate these options, we consider four routes **R1**, **R2**, **R3** and **R4** towards the same destination **D1**. Each of these routes encloses a set of QoS parameters. As illustrated in Figure 7, the route candidates' pool depends on the implemented option.

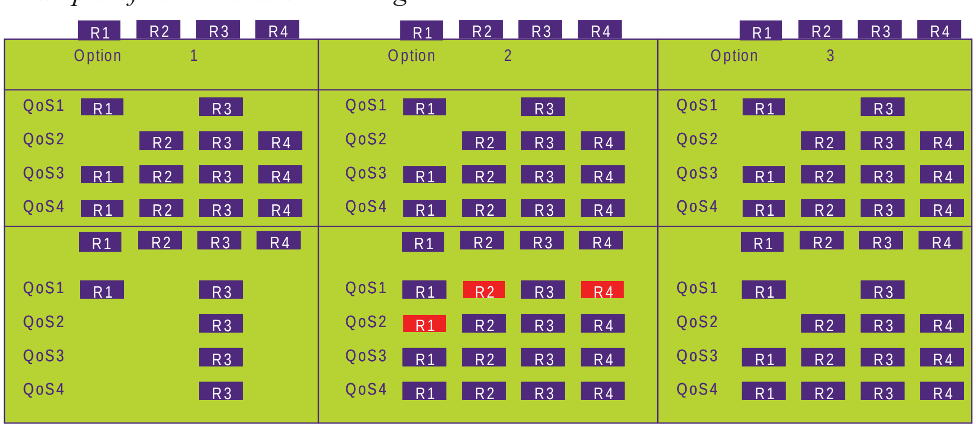
4 QOS-INFERRED PARALLEL INTERNET

4.1 Context

Building end-to-end QoS paths, for the purpose of QoS-guaranteed communications between end-users, is going a step further in the QoS process. The full description of customer-to-provider QoS agreements, and the way they are enforced, is outside the scope of this chapter. This section describes a scenario involving q-BGP and Meta-QoS-Class for the creation of QoS-Inferred Internet. Therefore, several Parallel Internets will be created. Each of these Parallel Internets will be engineered in such a way to be suitable to convey traffic belonging to IP services with similar/consistent QoS requirements.

We assume that Network Providers deploy similar classes of service because they are in general confronted with the same customers' requirements. These classes target to support applications, which have similar QoS constraints. There is no reason to consider that a Network

Figure 7. Example of route decision-making



Provider in France for instance would design a "Voice over IP" l-QC with short delay, low loss and small jitter while another one in Germany would have a completely different view. Therefore, constraints are implicitly imposed by applications to the network, independently of where the service is consumed or delivered.

As mentioned previously, the Meta-QoS-Class concept is an abstract concept. It is not a real l-QC provisioned in real networks. A Meta-QoS-Class is defined to serve dedicated services and can specify a set of boundaries for pertinent QoS performance attributes. The underlying philosophy of Meta-QoS-classes relies on the assumption that wherever end-users are connected they use similar applications in similar business contexts. Customers also experience the same QoS difficulties and are lead to express similar QoS requirements to their respective Network Providers.

4.2 Parallel Internet Framework

The idea behind Parallel Internet is to enable a Network Provider to offer its customers access to differentiated end-to-end services, where each differentiated service is related to a Meta-QoS-Class. It is envisaged that Network Providers throughout the Internet will implement a small number of well-known Meta-QoS-Classes and activate QoS-Enhanced BGP to interconnect their domains. Inter-domain QoS services are then created by constructing paths across those domains that support a particular Meta-QoS-Class. A set of parallel "internets" are deployed, each offering service levels associated with a specific Meta-QoS-Class. Reachability within the context of a given Meta-QoS-Class plane is ensured by q-BGP.

```
Parallel Internet = Meta-QoS-Class + QoS-Enhanced BGP
```

The guarantees associated with the service are restricted to qualitative services, although it is anticipated that the characteristics of each Meta-QoS-Class based service will be based on common application requirements. The main motivation behind the proposed rationale is to address the requirements of a large population of users, while keeping the network engineering as simple as possible by supporting relaxed service guarantees.

In order to establish the inter-domain Meta-QoS-Class planes the following steps need to take place:

- Establishment of provider-to-provider agreements prior to the exchange of traffic belonging to a certain Meta-QoS-Class.
- Identification of the traffic flows and q-BGP announcements that fall into a particular Meta-QoS-Class. For traffic flows this is achieved at a packet level by using the DSCP field and, in q-BGP, by means of a **QoS_NLRI** attribute. The identifiers could either be globally known or agreed between adjacent AS peers at the time of provider-to-provider establishment.
- Announcement of the network prefixes reachable within a given Meta-QoS-Class plane and associated QoS attributes. This is achieved dynamically through q-BGP.
- DSCP swapping of data packets at each domain's ingress and egress routers. Meta-QoS-Class traffic, when arriving at a domain, needs to be marked appropriately to receive the relevant local treatment. When exiting from a domain, Meta-QoS-Class traffic needs also to be remarked to the value agreed in the provider-to-provider agreement with the adjacent domain for that Meta-QoS-Class.

When traversing a set of ASes, the QoS treatment experienced by an IP datagram is consistent in the sense that packet treatment received in each domain conforms to the corresponding Meta-QoS-Class.

The resulting QoS Internet can be viewed as a set of parallel Internets or Meta-QoS-Class planes. Each plane consists of all the l-QCs bound according to the same Meta-QoS-Class. A Meta-QoS-Class plane can have holes and isolated domains because QoS capabilities do not cover all Internet domains. When an l-QC maps to several Meta-QoS-Classes, it belongs potentially to several planes. When a Network Provider contracts with another Network Provider based on the use of Meta-QoS-Classes, it simply adds a logical link to the corresponding Meta-QoS-Class plane. This is basically what current traditional inter-domain agreements mean for the existing Internet. Figure 8 depicts the physical layout of a fraction of the Internet, comprising four domains with full-mesh connectivity.

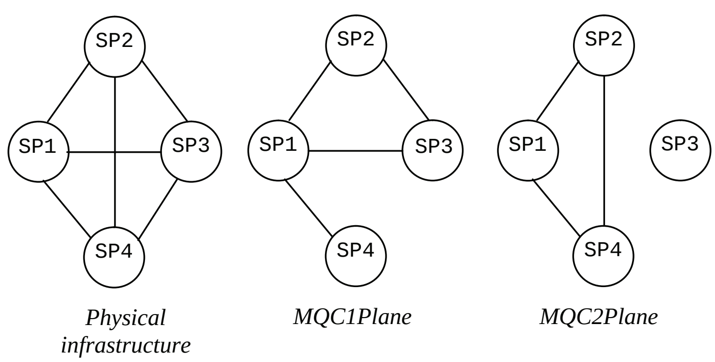
These four domains are involved in two Meta-QoS-Class planes. Let's focus on the **MQC2** plane. **SP1**, **SP2** and **SP4** have at least one compliant l-QC (**SP3** maybe has or not) for this Meta-QoS-Class. A bi-directional agreement exists between **SP1** and **SP2**, **SP1** and **SP4**, **SP2** and **SP4**.

Meta-QoS-Class brings a clear distinction between provider-to-provider and customer-to-provider QoS agreements. We expect a great deal of difference in dynamicity between the two. Most provider-to-provider agreements should have been negotiated, and should remain stable, before end-users can dynamically request end-to-end guarantees. Provider agreements do not directly map end users' needs, therefore the number of provider agreements is largely independent of the number of end-user requests and does not increase as dramatically as with SP chains.

For a global QoS-based Internet, this solution will work only if Meta-QoS-Class-based binding is largely accepted and becomes a current practice. This limitation is due to the nature of the service itself, and not to the use of Meta-QoS-Classes. Insofar as we target global services we are bound to provide QoS in as many SP domains as possible. However, any Meta-QoS-Class-enabled part of the Internet that forms a connected graph can be used for QoS communications, and be extended. Therefore, incremental deployment is possible, and leads to incremental benefits. For example, in Figure 8 **MQC2** plane, as soon as **SP3** connects to the **MQC2** plane it will be able to benefit from the **SP1**, **SP2** and **SP4** QoS capabilities.

The Internet as a split of different Meta-QoS-Class planes offers an ordered and simplified view of the Internet QoS capabilities. End-users can select the Meta-QoS-Class plane that is the closest to their needs, as long as there is a path available for the destination. One of the main outcomes of applying the Meta-QoS-Class concept is that it alleviates the complexity and the management burden of inter-domain relationships.

Figure 8. Different views of the same Internet part



5 CONCLUSION AND FUTURE TRENDS

This chapter introduced a novel framework to deliver QoS-enabled services across several domains managed by distinct Network Providers. This framework, known as Parallel Internet, abolishes the notion of administrative domains and ensures coherent and consistent treatment to be experienced by IP flows. Parallel Internet framework promotes the emergence of QoS-enabled end-to-end services. Also described within this paper, an implementation of Parallel Internet example based on the activation of an enriched version of BGP, called QoS-Enhanced BGP, together with a novel mean put at disposal of Network Provider to harmonize their mutual agreements: the Meta-QoS-Class. The concept of Meta-QoS-Class is a step forward to standardize provider-to-provider agreements and facilitate extending of services scopes beyond a single administrative domain.

The Parallel Internet framework is currently under investigation and enhancement in the context of IST AGAVE project. Issues related to availability and robustness of end-to-end services are taken into account. Moreover, the IST AGAVE project investigates how individual Network Providers contributes to the overall Parallel Internet, thus the concept of Network Planes. Network Planes is a concept that encloses both QoS-related and robustness considerations. Parallel Internet and Network Planes are in line with the ongoing initiatives (such as GENI) which ambition is to prepare the future Internet.

6 REFERENCES

Almes, G., Kalidindi, S., & Zekauskas, M. (1999). *A One-way Delay Metric for IPPM*, IETF, RFC 2679.

Almes, G., Kalidindi, S., & Zekauskas, M. (1999). *A One-way Packet Loss Metric for IPPM*, IETF, RFC 2680.

Atkinson, R., & Floyd, S. (2004). IAB Concerns & Recommendations Regarding Internet Research & Evolution, IETF, RFC 3869.

Atkinson, R. (Ed.). (2004). IAB Concerns and Recommendations Regarding Internet Research and Evolution, IETF, RFC 3869.

Bates, T., Chandra, R., Katz, D., & Rekhter, Y. (2007). *Multiprotocol Extensions for BGP-4*, IETF, RFC 4760.

Blake, S. (Ed.). (1998). An Architecture for Differentiated Services, IETF, RFC 2475.

Boucadair, M. (2005). *QoS-Enhanced Border Gateway Protocol*, draft-boucadair-qos-bgp-spec-01.txt.

Boucadair, M. (Ed.). (2006). *Parallel Internets Framework*, IST AGAVE project, deliverable D1.1, from http://www.ist-agave.org/results/D1.1-final-public.pdf.

Chandra, R., & Scudder, J. (2002). *Capabilities Advertisement with BGP-4*, IETF, RFC 3392.

Deleuze, G., & Guattari, F. (1996). *What Is Philosophy?* Columbia University Press ISBN: 0231079893.

Demichelis, C., & Chimento, P. (2002). *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)*, IETF, RFC 3393.

Huston, G. (2000). *Next Steps for the IP QoS Architecture*, IETF, RFC2990.

Levis, P., Asgari, H., & Trimintzios, P. (2004). Consideration on Inter-domain QoS and Traffic Engineering issues Through a Utopian Approach. *SAPIR-2004 workshop of ICT-2004*, (C) Springer-Verlag.

Nichols, K., & Carpenter, B. (2001). *Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification*, IETF, RFC 3086.

Rekhter, Y., Li, T., & Hares, S. (2006). *A Border Gateway Protocol 4 (BGP-4)*, IETF, RFC 4271.

Saltzer, J. H., Reed, D. P., & Clark, D. D. (1984). End-To-End Arguments in System Design. *ACM Transactions in Computer Systems, 2*(4), 277-288.

7 KEY TERMS

Domain: A network infrastructure composed of one or several Autonomous Systems managed by a single administrative entity.

Extended QC: A QoS transfer capability provided using both the local domain and neighbour domains. An extended QC is provided by combining the local QC of local domain with the ones of adjacent domains. The topological scope of an extended QC extends the boundaries of local domain.

Inter-Domain QoS Delivery Solution: Is used to denote an inter-domain system that aims at offering inter-domain QoS services.

IP Connectivity Service: Transfer capability characterized by a **(Destination, D, J, L)** tuple where Destination is a group of IP addresses and **(D, J, L)** is the QoS performance to get to Destination.

Local-QoS-Class (l-QC): A QoS transfer capability across a single domain, characterized by a set of QoS performance parameters denoted by **(D, J, L)**. From DiffServ perspective, an l-QC is an occurrence of a Per Domain Behavior.

L-QC Binding: Two l-QCs from two neighboring domains are bound together once the two providers have agreed to transfer traffic from one l-QC to the other.

L-QC Thread: Chain of neighboring bound l-QCs.

Meta-QoS-Class (MQC): An MQC provides the limits of the QoS parameter values that two l-QCs must respect in order to be bound together. An MQC is used as a label that certifies the support of a set of applications that bear similar network QoS requirements.

QoS-Enhanced BGP (q-BGP): An enhanced BGP that takes into account QoS information it carries in its messages as an input to its route selection process.

"QoS Service"-Related Attributes: Denotes dedicated q-BGP attributes for the usage of a given QoS service;

QoS-Related Information: Can be expressed in terms of one-way delay, inter-packet delay variation, loss rate, DSCP marking, or a combination of these parameters;

Service Provider (SP): An entity that provides Internet connectivity. We assume that an SP owns and administers an IP network called a domain. Sometimes simply referred to as provider.

SP Chain: The chain of Service Providers whose domains are used to convey packets for a given IP connectivity service.

Chapter XLIV
The Playout Control Management: An Issue for the IP Telephony Service Providers

Mirko Luca Lobina
Università di Cagliari, Italy

Luigi Atzori
Università di Cagliari, Italy

Fabrizio Boi
Università di Cagliari, Italy

ABSTRACT

IP Telephony provides a way for an enterprise to extend consistent communication services to all employees, whether they are in main campus locations, at branch offices, or working remotely, also with a mobile phone. IP Telephony transmits voice communications over a network using open standard-based Internet protocols. This is both the strength and weakness of IP Telephony as the involved basic transport protocols (RTP, UDP, and IP) are not able to natively guarantee the required application quality of service (QoS). From the point of view of an IP Telephony Service Provider this definitely means possible waste of clients and money. Specifically the problem is at two different levels: i) in some countries, where long distance and particularly international call tariffs are high, perhaps due to a lack of competition or due to cross subsidies to other services, the major opportunity for IP Telephony Service Providers is for price arbitrage. This means working on diffusion of an acceptable service, although not at high quality levels; ii) in other countries, where different IP Telephony Service Providers already exist, the problem is competition for offering the best possible quality. The main idea behind this chapter is to analyze specifically the state of the art playout control strategies with the following aims: i) propose

the reader the technical state of the art playout control management and planning strategies (overview of basic KPIs for IP Telephony); ii) compare the strategies IP Telephony Service Provider can choose with the aim of saving money and offering a better quality of service; iii) introduce also the state of the art quality index for IP Telephony, that is a set of algorithms for taking into account as many factors as possible to evaluate the service quality; iv) provide the reader with examples on some economic scenarios of IP Telephony.

INTRODUCTION

The combination of IP and a telephonic service gives IP Telephony. IP Telephony implements services like sending/receiving/management of voice and data-voice, between two or more users in a real time fashion over an already existing IP channel. The basic frameworks for implementing an IP Telephony solution are ITU-T H.323 (Rec. H.323, 2006) and SIP (RFC 3261, 2002). The ITU-T H.323 is a recommendation that defines the protocols to provide audio-visual communication sessions on any packet network, involving both the management of the call and the speech packet transport, including the speech coding, (see Figure 1).

The H.323 standard specifies four kinds of components, providing a point-to-point and point-to-multipoint IP Telephony service (see also Figure 2 from packetizer.com):

- **Terminals:** *A personal computer (PC) or a stand-alone device, running an H.323 and the multimedia applications*

Figure 1. The ITU-T H.323

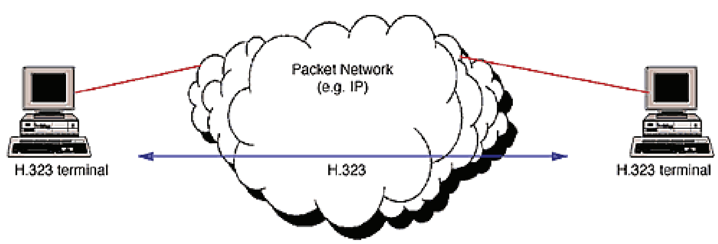
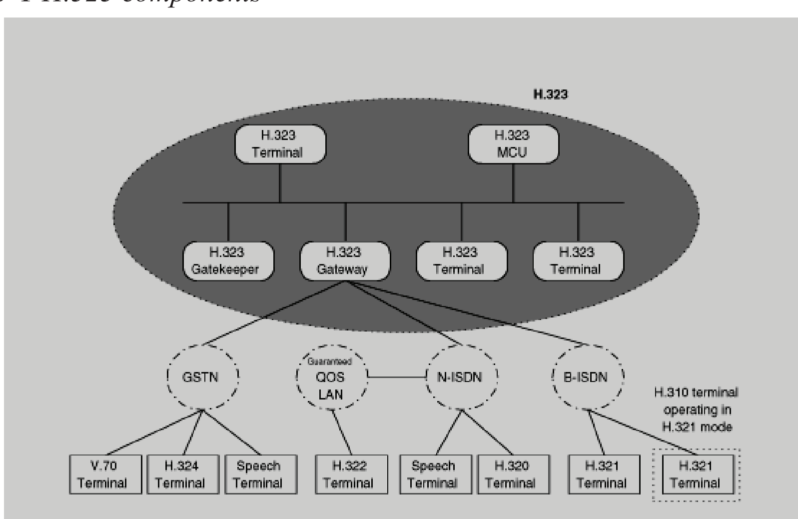


Figure 2. The ITU-T H.323 components



- **Gateways:** *Providing of connectivity between an H.323 network and a non–H.323 network*
- **Gatekeepers:** *The focal point for all calls within the H.323 network*
- **Multipoint control units (MCUs):** *Support for conferences of three or more H.323 terminals*

All these components and the related operations are detailed in the ITU-T documents. The ITU-T H.323 is an umbrella recommendation that comprises directly, or through a reference, all the standards needed for coding, transporting and managing IP Telephony sessions. As an example, coding operations are referenced in ITU-T H.323 documents but described in detail in other recommendations such as ITU-T G.723 (Rec. G.723.1, 2006) and G.729 (Rec. G.729, 2007).

Furthermore, H.323 is part of the H.32X family, which comprises different recommendations (see Tab.1).

On the contrary, the Session Initiation Protocol (SIP) (RFC 3261, 2002) is only an application-layer control (signaling) protocol for managing sessions with one or more participants. It can be used to create two-party, multi-party, or multicast sessions that include Internet telephone calls, multimedia distribution, and multimedia conferences. SIP clients use TCP or UDP to connect to SIP servers and other SIP endpoints. SIP is primarily used in setting up voice/video calls. However, it can be used in any application where session initiation is a requirement.

Table 1. The ITU-T H.32X family

H.32X	*Packet Network*
H.322	LANs that provide guaranteed QoS
H.324	SCN
H.320	integrated services digital networks (ISDN)
H.321 and H.310	broadband integrated services digital networks (B–ISDN)

A comparison between H.323 and SIP is not purposeful and possible without considering technical details about specific features. More in general, H.323 is the more mature of the two, but it is not very flexible. On the contrary, SIP is currently less defined, but has greater scalability which could ease Internet application integration. In addition, both standards are continuously updated and there is much work in progress. For this reason we suggest the reader to refer always to the latest versions and related documents.

Independently from the implementation, IP Telephony can be identified in several economic scenarios, three of which are described as follows. The *first* could be realistic in some countries, where long distance and international call tariffs are high, both for incoming and outcoming calls. Users may choose IP Telephony offerings to avoid the expensive tariffs of the classic telephony service and the international accounting rate system. The presence of IP Telephony could lead to a reduction of the normal call tariffs (e.g., some experiences have shown evidences of reduction in international tariffs due to the use of "compressed" or "packetized voice" on international IP circuits). A *second* group of countries is identified where call prices are already falling, for instance as part of a liberalized market. The evidences have shown that, in this context, IP Telephony can be a good antagonist to classic telephony in long and international calls. Also in this case, this generally results in lower prices for the user. A *third* group is composed of countries where call prices are generally low, as a result of intense market competition. Here the opportunities for IP Telephony service providers are significantly reduced in terms of prices, although it may still offer carriers opportunities mainly related to new value-added applications.

Given these, first of all, a general background of IP Telephony is provided: i) most important Key Performance Indicators (KPIs); ii) quality evaluation; iii) general introduction to *playout* buffering. The second step is a focus on the main theme of the paper (*playout* buffering): description and comparison of the most diffused and important strategies. Then, several future trends are illustrated and finally some conclusions are drawn.

BACKGROUND

The three main network-related KPIs of IP Telephony are: delay, loss and jitter. In the following a brief description of such KPIs is offered, trying where possible to relate them to a subjective quality scale. Delay (also named latency) is the time delay incurred in speech by the IP Telephony system. Delay is typically measured in milliseconds (*msec*) from the moment that the speaker produces a speech until the listener hears that speech. This is termed as "mouth-to-ear" delay or the "one-way" delay. In the traditional Public Switched Telephone Network, the round-trip latency for domestic calls is virtually always under 150 milliseconds. A set of delay specifications from Cisco is provided in Table 2, while subjective (or perceived) quality is directly linked to the delay as appears in Figure 3.

The loss event is always related to network. An IP network presents some short-term correlation features, that is if the n*th* packet is lost also the (n+1)*th* packet could be lost. This phenomenon causes the loss burstiness. Apart from the overall percentage of lost packets, the burstiness is important, in particular, for a correct estimation of the speech quality. In fact, approximating bursty losses as they would be random implies an overestimation of speech quality degradation. The loss is usually modeled by means of a 2-state Gilbert model (Jiang, 2000). Usually the threshold of overall packet loss rate for a good quality is fixed to 5%. This means that for greater values of overall loss the perceived quality is unacceptable.

Table 2. The ITU-T H.32X family

Delay (msec)	***Description***
0-150	Acceptable for most user applications
150-400	Acceptable provided that administrators are aware of the transmission time and the impact it has on the transmission quality of user applications
Above 400	Unacceptable for general network planning purposes. However, it is recognized that in some exceptional cases this limit is exceeded

Figure 3. Delay and perceived quality

The third KPI, the jitter, is caused by the high variability of the network component of the one-way delay. It could be related to some time-variant network factors, such as network congestion, improper queuing, or configuration errors. The result is that the destination receives a flow of packets with irregular inter-packet delays. This point is better explained in the following. At the sending side the instants of generation of the speech packets are *k*T*, where *k* is the number of the emitted transport packet (*0…n*) while *T* depends on the number and size of the speech frames conveyed in every transport packet (e.g., for the ITU-T G.729 speech codec, usually T = 20 ms, corresponding to 2 speech frames of 10 ms). When these packets are sent to the network their order can change due to network variable conditions. This causes an alteration in the order of arrival of transport packet: as an example, packet *X* can arrive later than *X+1*. Thus, a buffer at the receiver is needed to reconstruct the right order and guarantee a good decoding. The usage of this buffer is at the base of a *playout* control mechanism. The setting of the size of the buffer is a key operation in IP Telephony as it directly affects the packet loss rate and the conversational interactivity, thus influencing the quality perceived by the end-user. Buffer adjustments may be introduced to avoid packet losses at all, to keep the packet loss under a desired threshold, or to have the optimal balance between loss and delay on the basis of a reference quality model. Not only is the absolute size value important, but when it is modified: the common attitude is to avoid buffer adjustments during the periods of speech.

A MOS score can be retrieved by means of an objective quality assessment methodology applied to the described KPIs, the typology of network and methods of accessing the IP Telephony service. Usually, objective quality assessment methodologies make use of tools called opinion models (Takahashi, 2004). These are aimed at evaluating the end-user perceived quality in two-way voice communication taking into account all the impairments introduced by the signal processing modules and the underlying network. One of the most well known and documented assessment methodology is the E-Model. This is an opinion model initially developed and standardized by the ITU-T and subsequently also adopted by the ETSI and the TIA, becoming the most widely used tool for objective quality assessment of the conversational quality.

The main feature of the E-Model is the capacity of revealing the underlying causes of speech quality problems by combining a set of elements (e.g., speech low bit rate coding; delay and loss distribution; frame erasure distribution; loss concealment technique; architecture choices such as *playout* control strategy, packet and codec frame sizes) in a formula that produces an overall index, the R Factor (to be mapped in a MOS score with the formula $MOS = 1 + 0.035R + 7x10\text{-}6R\ (R\text{-}60)\ (100\text{-}R)$). All these elements are expressed by four factors, summed to obtain the R Factor, $R = 100 - I_s - I_d - I_{e,eff} + A$ (Rec. G107). From left to right, the four terms are the simultaneous impairment, delay impairment, equipment impairment, and the advantage factor. I_s, the simultaneous impairment, comprises the distortions introduced by the circuit-switched part of the transmission path and is frequently set to the default value of 6.8 (Rec. G107). I_d , delay impairment, measures the degradation related to the mouth-to-ear delays encountered along the transmission path. The analytical expression of $I_d = I_d(d)$ has been experimentally derived and can be found in (Cole, 2001) under the hypothesis of random losses. The third element $I_{e,eff}$ represents the impairments associated with the signal distortion, caused by the end-to-end packet loss rate (taking also into account the typology of speech codec used during the IP Telephony session). The end-to-end packet loss rate comprises both the packets that have been lost during transmission, after the application of the FEC loss recovery (if used), and those that arrive correctly at the receiver but are too late to be played out. Rec. G.113 (2001) provides the $I_{e,eff} = I_{e,eff}(LOSS)$ values for some configurations, including: a fixed loss rate; the number of frames inserted in a transport packet; the distribution of the packet losses; the sensitivity of the used codec to data frame losses and the concealment algorithm. Finally, the advantage factor A raises the level of conversational quality when the end-user may accept some decrease in quality for access advantage, such as mobility. A list of values for the A factor is provided in (Rec. G107).

In general, the E-Model represents a very powerful tool in the process of network planning and troubleshooting, even if its accuracy and validity are often argued (especially its additive property). The description of the E-Model is purposely brief to give the reader only an overview of the methodology and a basis for the understanding of the *playout* quality-based strategies described in the following section.

MAIN FOCUS OF THE CHAPTER

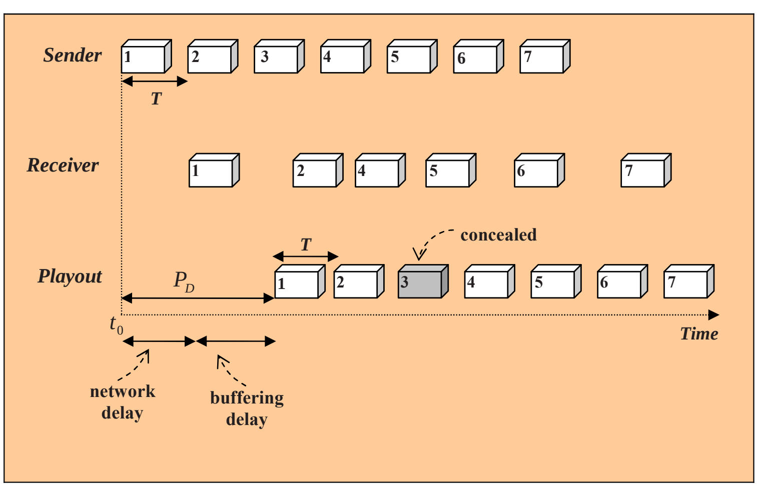
The jitter phenomenon is sketched in Figure 4 in the middle row of packets (note that some packets may also be missing due to network losses, as happened to packet 3). The jitter is quite harmful since the decoder, to avoid the degradation of the listening quality, should play out the data smoothly (that is, with the same interval T of the sender side). To this aim, a *playout* buffer is used at the receive-side to compensate the transmission jitter with the addition of additional delay in the end-to-end chain. The intent is to re-obtain a sequence of *playout* instants uniformly spaced, as illustrated in the bottom row of the figure (i.e., the distance between each couple of consecutive packets should be the same as during voice coding).

Accordingly, the delay between the departure and the decoding, called end-to-end delay or *playout* delay, is kept equal for every packet (at least for a group of packets, usually corresponding to an entire talkspurt) and given by the sum of the buffering and the network delays, which instead are both random. The setting of the *playout* delay for the first packet is shown in the figure. Usually, the receiving end discards the packets that arrive later than their scheduled instants (*playout* instants) and tries to conceal these losses usually predicting the missing parameters from the previously received packets. Less often, it is not discarded, but the previous packet content is stretched till the arrival of the late packet so as not to stop the play out. The *playout* algorithms are devoted to setting the end-to-end delay, which is changed during the streaming session as described in the following.

The problem of jitter compensation has been addressed by using several approaches since the early 80's. Montgomery (1983) and Barberis (1980) are some of the first works that propose the introduction of additional delay to solve the problem of *playout* synchronization between sender and receiver. Since then, many other strategies have been investigated. In the following, we propose a classification that is summarized in Figure 5. At first, two groups are introduced, that is, those of fixed and adaptive approaches. According to the techniques belonging to the first group, the end-to-end delay is kept constant for all voice packets in a session (Gruber, 1981; Montgomery, 1983). Such an approach is inefficient, since a fixed configuration is not reliable with the temporal variability of the network behavior. Differently, techniques in the second group work adapting the delay to the changing network conditions, preventing from using high delays during low congestion conditions and vice versa. In this case, the network behavior is monitored during the streaming session and the buffer size adjusted accordingly.

One of the main features of the adaptive algorithms is that *playout* buffer is adjusted. Intra-talkspurt techniques modify the end-to-end delay

Figure 4. Playout buffering to compensate the transmission jitter



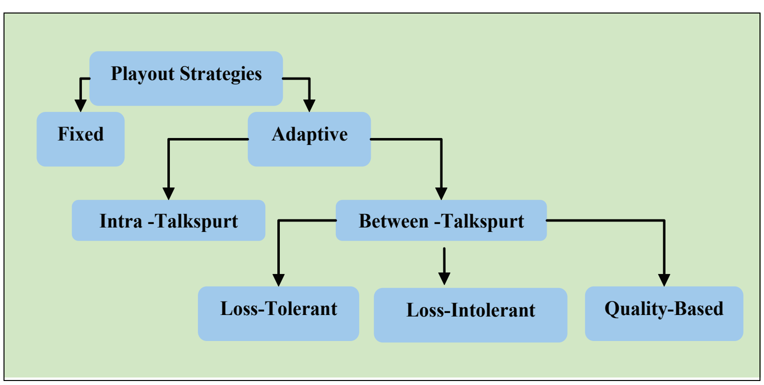
independently from the silence periods, using some strategies of compression and extension of the waveform. On the other hand, between-talkspurt methods act during the intervals of silence (Rec. 107, 2003). The latter methods are more frequently used since they don't require any signal processing techniques to change the length of the speech. On the contrary, they are not able to take into account high-delay spikes within talkspurts, resulting occasionally in a burst of losses and audible quality degradation. Among the adaptive between-talkspurt techniques, some estimate the network delay statistics and set the *playout* delay so that only a small fraction of packets arrive late. This approach, which we call loss-intolerant, does not take into account any loss concealment, resulting in an overestimation of the required the *playout* delay. Indeed, IP telephony applications can tolerate or conceal a certain amount of late packets; accordingly, some *playout* techniques perform a controlled tradeoff between the packet loss and the delay. A percentage of packet loss is then allowed and the *playout* delay is set so as to reach this "target value". We refer to these as loss-tolerant techniques.

Another category of algorithms is named quality-based; these are driven by the maximization of a metric linked to the end-user perceived quality. These algorithms appeared in the literature only few years ago, and have been triggered by the advancements reached by the standardization committees working on the development of voice QoS evaluation tools within ITU (Rec. G.107, 2003; Rec. G.114, 2000), ETSI (Clark, 2001; TIPHON TS 101329-5, 2001), and ANSI (T1A1.7/98-031, 1998) organizations.

In this chapter, we focus mainly on the between-talkspurts strategies, more frequently used for computational complexity reasons. In the following the loss-intolerant, loss-tolerant and quality-based strategies are presented together with a brief list of pro/contro factors. The idea is not to give a detailed explanation but only the indication of the state of art within the indicated categories with a minimum of suggestions to the reader to perform an implementation choice. The first category is the *Loss-intolerant*. Such techniques work between talkspurts; it means that changes in the *playout* delay are introduced only during silence periods not requiring the decoder to perform waveform adjustments. The underlying idea is that packet losses are undesired in a voice packet communication and should be avoided even at the expense of high end-to-end delays (not considering an eventual FEC system). One of the first works based on this approach is by Ramjee et al. (Ramjee, 1994), who proposed the use of a *playout* delay high enough to expect only a small fraction of packet losses.

Essentially, the *playout* delay is estimated making use of a linear recursive filter characterized by a fixed weighting factor, α. But, choosing a fixed

Figure 5. A classification of playout strategies



α may be inappropriate under typical present-day Internet conditions. For this reason, several variants of this algorithm have been proposed. The α-Adaptive technique (Kansal, 2001) generalizes the filtering method in (Ramjee, 1994) by using several values of α. The adaptation is performed by looking at the resulting packet loss rate during the past in case of using the current α value minus/plus Δ; the correct value of α is then chosen. Differently, in (DeLeon, 1999), the application of a Normalized Least Mean Square (NLMS) active predictor is proposed. This is used to estimate the network delay for each packet from the previous N ones. Narbutt (2004) is the most recent work based on the loss-intolerant approach, that is based on a dynamic adjustment of α.

The main advantage of the loss-intolerant approach is to be straightforward and then simple to be implemented. We generally believe that this approach has to be employed when the user is aware of the network condition which should be stable so as not to have too high end-to-end delays. *Loss-intolerant* algorithms in last section estimate an average network delay and use it to set the *playout* delay so that the fraction of late packets is kept very small. However, IP telephony applications can tolerate or conceal a small amount of late packets after the *playout* buffer. On the basis of this observation, some strategies have been devised. They are called *Loss-tolerant*. The Concord algorithm in (Sreenan, 2000) is maybe one of the most known strategies. It fixes two thresholds for the maximum late packet percentage and maximum acceptable delay. The strategy makes use of a histogram of the network delay observed for past packets; from the histogram, an approximated and sampled version of the packet delay distribution (PDD) is computed. By using such PDD, the algorithm set the *playout* delay so that the two constraints on the maximum late packets percentage and maximum acceptable delay are satisfied.

Other good examples of Loss-tolerant strategies are presented in (Fujimoto, 2001; Moon, 1998), and basically differ from the concord in a way the PDD is updated and considered. The authors suggest also a reading of the work presented in (Rosemberg, 2000), whom main contribution consists in the evaluation of the interactions between FEC (Forward Error Correction) and *playout* buffering. In particular, the authors evaluate the repair capability of the FEC scheme and its influence on both the total loss percentage and the network delay. The last category is the *quality-based* approach. The E-MOS algorithm in (Fujimoto, 2002) is the one of the first *playout* buffer techniques that tries the way of using quality metrics. This work relies on the studies by Savolaine in (Savolaine, 2001), who provides the experimental relationships between MOS (Rec. P.800, 1996) and the *playout* delay at different packet loss rates for the ITU-T G.711 speech codec.

The authors strongly suggest also the work presented in (Savolaine, 2001). In (Boutremans, 2003), the joint problem of FEC and *playout* buffering is addressed. The quality is evaluated making use of the ITU-T E-Model with some additional elaborations. One of the main contributions of the work in (Sun, 2004) is the evaluation of the listening quality for modern speech codecs, such as G.723.1, G.729, AMR and iLBC, to extend the applicability of the E-Model. To this purpose, the PESQ-LQ index (Rix, 2003), an improvement of the PESQ index, is used to assess the quality of the speech with different codecs and at varying packet losses, modeled as a Bernoulli process. The results are then converted into the R Factor and the resulting points are interpolated to achieve a three-parameter logarithmic expression of the impairment factor respect to the packet loss rate. These expressions are then used for *playout* buffer setting, minimizing the sum of delay and equipment impairment factors. Atzori (2004) proposes the use of simplified expressions of the equipment and delay impairment factors resulting from the study of Cole and Rosenbluth in (Cole, 2001). Note that, the used simplifications are valid when: the

ITU-T G.729-A speech codec is used; the losses are random. The works proposed in (Boutremans, 2003; Sun, 2004; Rix, 2003; Atzori, 2004) are all based on the ITU-T E-Model, which represents the most accurate tool for the required quality evaluation (some considerations about future works for ITU-T E-Model extension are proposed in the following Section).

In the following we propose a comparison of strategies performance between different categories of frameworks. Within the Loss-intolerant algorithms, we have observed that the original liner predictor (Ramjee, 1994) has the tendency to be conservative during the normal operation mode, with many overestimates and their propagation over the time. On the contrary, the α-adaptive algorithm (Kansal, 2001) follows the delay variations in a way overestimations are quite avoided. Better results have been obtained with (Narbutt, 2004), which works dynamically adapting α according to the variation of the delay. From this analysis, it results that the setting of the linear filter parameter is crucial in the estimation and should quickly change to follow the network behaviour. A further comparison between (Narbutt, 2004; Sreenan, 2000) (implemented with and without the aging of the PDD) and (Atzori, 2006) shows that in case of high jitter the best performance have often been obtained with the (Narbutt, 2004) algorithm, which provided the best compromise in terms of loss/delay ratio. It can be also noted that (Sreenan, 2000) benefits from the aging procedure in the PDD, which allows for lower *playout* delays at medium-low packet loss rates.

The quality-based algorithm in (Atzori, 2006) provided performance between the Loss-tolerant/intolerant, showing that maximizing the user-perceived quality doesn't mean minimizing the loss-delay ratio. In the following we present also the comparison in terms of R Factor and MOS (see Table 3). We note that (Atzori, 2006) achieves the best result in terms of R Factor (and MOS), while the second place goes to (Narbutt, 2004) and the third to (Sreenan, 2000) with aging. The first position for (Atzori, 2006) is not surprising since this strategy is born to maximise quality. As already underlined this does not imply the minimization of the *playout* delay. The second position for (Narbutt, 2004) is really surprising, considering also the basic simplicity of the strategy itself. To provide the reader an idea of the computational complexity we present also Table 4, where M is number of bins in histogram and Z is number of considered *playout* delay values. Again we note that (Atzori, 2006) is the more complex strategy as compared to the others.

Table 3. Comparison based on the E-Model quality model

Strategy	Quality	
Codec	G.729	G.723.1
Ramjee, 94	53.3 (2.71)	55.9 (2.84)
Kansal, '01	57.2 (2.82)	59.5 (3.02)
Narbutt, '04	67.6 (3.48)	67.7 (3.50)
Sreenan, 00	57.2 (2.82)	59.5 (3.02)
Sreenan, 00-aging	64.1 (3.32)	65.2 (3.38)
Atzori, 06	70.1 (3.62)	70.5 (3.76)

Table 4. Comparison Analysis of the computational complexity

Strategy	Complexity during talkspurt	Complexity during silence
Ramjee, 94	$O(1)$	$O(1)$
Kansal, '01	$O(1)$	$O(1)$
Narbutt, '04	$O(1)$	$O(1)$
Sreenan, 00	$O(1)$	$O(M)$
Sreenan, 00-aging	$O(M)$	$O(M)$
Atzori, 06	$O(Z)$	$O(Z)$

FUTURE TRENDS

Here we present some future trends about both the *playout* control strategies and some fields that have been presented/underlined in this work. The first point is about the quality models used in the Quality-based frameworks. Obviously the rule "the better the model the better the estimation" is valid. This means that when evaluating if applying such techniques the reader should consider using the best "state of the art quality" strategies in the market/research places. As an example the presented ITU-T E-Model shows some drawbacks: the overall additivity property of the model is applicable only to a certain extent; many assumptions on the configuration of the network and service have to be introduced to make the model applicable in a *playout* buffering context; experiments with low bit rate voice codecs are missing in case of bursty losses.

As to the burstiness of packet losses, this is indeed an important issue since several studies have shown the burstiness of packet losses in the Internet (Jiang, 2000). Dealing with bursty losses as if these were random would introduce many issues: at a given total loss rate, the subjective impact of isolated losses and grouped losses is quite different. All these aspects have to be taken into account when using a quality-based approach not to use the tool inappropriately. The second consideration regards the crossing use of video/voice, as the future trends for real time services show the integration of different services. The idea here is the finding of a unique value of *playout* buffering that could be optimum (or at least acceptable) both for audio and video, independently from the used category (Loss-tolerant/intolerant or Quality-based).

The other future trend is about the introduction of right extensions to a quality model for the usage/application on a mobile context. In fact many vendors are introducing and using the IP Telephony for mobile through an UMTS/GPRS service.

CONCLUSION

The jitter is an innate and damaging factor in IP Telephony. It is caused by the high variability of the network component of the delay, caused by some underlying time-variant network factors, such as network congestion, improper queuing, or configuration errors. The jitter causes irregular inter-packet delays at the destination side and this is unacceptable for a correct decoding and *playout*. To solve this problem, a *playout* control mechanism is used at the receiver side to compensate the jitter at the expense of an additional delay (introduced by the buffer). The intent is to obtain a new sequence of transport packets, uniformly spaced and decoded. The logic is quite simple: stop the received packets in the buffer for a certain interval of time in order to reconstruct the exact order (identical to the sender side) of the transport packets. Usually a *playout* mechanism sets the end-to-end delay, which includes also the *playout* delay (i.e., the dimension of the buffer). This setting is performed several times during an IP Telephony session and several *playout* control strategies exist that perform this task with different approaches.

This chapter has proposed a classification of the state of the art *playout* control strategies with the following aims: i) propose the reader the technical state of the art *playout* control management and planning strategies; ii) help the reader to choose the most desirable framework by comparing the strategies; iii) introduce the ITU-T E-Model quality index for IP Telephony; iv) provide the reader with examples on some economic scenarios of IP Telephony; v) present some future trends in IP Telephony.

REFERENCES

Atzori, L., & Lobina, M. L. (2004). Speech *playout* buffering based on a simplified version of the

ITU-T E-Model. *IEEE Signal Processing Letter, 11*(3), 382-385.

Atzori, L., Lobina M. L., & Corona, M. (2006). *Playout* Buffering of Speech Packets based on a Quality Maximization Approach. *IEEE Transaction on Multimedia 8*(2), 420-426.

Barberis, G., & Pazzaglia, D. (1980). Analysis, design and buffer sizing of a packet voice receiver. *IEEE Transactions on Communications,* 217–227.

Boutremans, C., & Le Boudec, J. Y. (2003). Adaptive *playout* buffer and FEC adjustment for Internet Telephony. *In Proceedings of 22th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)* (pp. 652-662).

Clark, A. (2001). Extensions to the E Model to incorporate the effects of time varying packet loss and recency. *ETSI T1A1.1/2001-037: Telchemy Incorporated*.

Cole, R., & Rosenbluth, J. (2001). Voice over IP performance monitoring. ACM Computer Communication Review, *31*(2), 9-24.

DeLeon, P., & Sreenan, C. J. (1999). An adaptive predictor for media *playout* buffering. In ICASSP (1999), *Acoustics, Speech, and Signal IEEE Processing*, *6*, 3097-3100.

Fujimoto, K., Ata, S., & Murata, M. (2001). *Playout* control for streaming applications by statistical delay analysis. *In Proceedings of IEEE International Conference on Communications (ICC 2001),8*, 2337-2342.

Fujimoto, K., Ata, S., & Murata, M. (2002). Adaptive *playout* buffer algorithm for enhancing perceived quality of streaming applications. *In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM 02)* (pp. 2463-2469), Taipei, Taiwan.

Gruber, J. G. (1981). Delay related issues in integrated voice and data networks. *IEEE Transactions on Communications, 29*, 786–800.

Jiang, W., & Schulzrinne, H. (2000). Modeling of packet loss and delay and their effect on real-time multimedia service quality. *In Proceedings of International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2000).*

Kansal, A., & Karandikar, A. (2001). *Jitter-free audio playout over best effort packet networks.* ATM Forum International Symposium, New Delhi, India.

Montgomery, W. (1983). Techniques for Packet Voice Synchronization. *IEEE Journal on Selected Areas in Communications*, *1*(6), 1022–1028.

Moon, S. B., Kurose, J., & Towsley, D. (1998). Packet audio *playout* delay adjustment: performance bounds and algorithms. *Multimedia Systems*, *6*, 17–28.

Narbutt, M., & Murphy, L. (2004). Improving voice over IP subjective call quality. *IEEE Communications Letters, 8*(5), 308-310.

Ramjee, R., Kurose, J., Towsley, D., & Schulzrinne, H. (1994). Adaptive *playout* mechanisms for packetized audio applications in wide-area networks. In INFOCOM (13th IEEE Proceedings), *Networking for Global Communications, 2,* 680-688. Toronto, Canada.

Rec. G.107 (2003). The E-Model, a computational model for use in transmission planning. *ITU-T Recommendation*.

Rec. G.113 (2001). Transmission impairments due to speech processing. *ITU-T Recommendation.*

Rec. G.114 (2000). One-way transmission time. *ITU-T Recommendation.*

Rec. G.723.1 (2006). Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. *ITU-T Recommendation.*

Rec. G.729 (2007). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). *ITU-T Recommendation*.

Rec. H.323 (2006). Packet-based multimedia communications systems. *ITU-T Recommendation.*

Rec. P.800 (1996). Methods for objective and subjective assessment of quality. Rec *ITU-T Recommendation.*

RFC 3261 (2002). SIP: Session Initiation Protocol. *IETF Standards Track.*

Rix, A. W. (2003). Comparison between subjective listening quality and P.862 PESQ score. *In Proceedings of Measurement of Speech and Audio Quality in Networks (MESAQIN 2003)* (pp. 17-25).

Rosenberg, J., Qiu, L., & Schulzrinne, H. (2000). Integrating packet FEC into adaptive voice *playout* buffer algorithms on the Internet. *In Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000)* (pp. 1705 - 1714), Tel Aviv, Israel.

Savolaine, C. (2001). QoS/VoIP overview. *In Proceedings of IEEE Communications Quality & Reliability International Workshop (CQR 2001).*

Sreenan, C. J., Chen, J. C., Agrawal, P., & Narendran, B. (2000). Delay reduction techniques for *playout* buffering. *IEEE Transactions on Multimedia*, *2*(2), 88-100.

Sun, L., & Ifeachor, E. (2004). New models for perceived voice quality prediction and their applications in *playout* buffer optimization for VoIP networks. *In Proceedings of IEEE International Conference on Communications (ICC 2004),. 3,* 1478-1483.

T1A1.7/98-031 (1998). Testing the quality of connections having time varying impairments. Committee contribution *ANSI & AT&T.*

Takahashi, A., Yoshino, H., & Kitawaki, N. (2004). Perceptual QoS assessment technologies for VoIP. *IEEE Communications Magazine, 42*(7), 28-34.

TIPHON TS 101329-5 (2001). QoS measurement methodologies. *ETSI Technical Standard: Annex E.*

KEY TERMS

European Telecommunications Standards Institute (ETSI): The European Telecommunications Standards Institute (ETSI) produces globally-applicable standards for Information and Communications Technologies (ICT), including fixed, mobile, radio, converged, broadcast and internet technologies. ETSI is a not-for-profit organization with almost 700 ETSI member organizations drawn from 60 countries world-wide (from www.etsi.org).

International Telecommunication Unit (ITU): ITU is the leading United Nations agency for information and communication technologies. As the global focal point for governments and the private sector, ITU's role in helping the world communicate spans 3 core sectors: radiocommunication, standardization and development. ITU also organizes TELECOM events and was the lead organizing agency of the World Summit on the Information Society. ITU is based in Geneva, Switzerland, and its membership includes 191 Member States and more than 700 Sector Members and Associates (from ww.itu.int).

Internet Protocol (IP): The Internet Protocol (IP) is a network-layer (Layer 3) protocol that contains addressing information and some control information that enables packets to be routed. IP is documented in RFC 791 and is the primary network-layer protocol in the Internet protocol suite. Along with the Transmission Control Protocol (TCP), IP represents the heart of the Internet protocols. IP has two primary responsibilities:

providing connectionless, best-effort delivery of datagrams through an internetwork; and providing fragmentation and reassembly of datagrams to support data links with different maximum-transmission unit (MTU) sizes networks (from www.cisco.com documentation).

ITU-T G.711. G.711: Is an ITU-T standard for audio companding. It is primarily used in telephony. The standard was released for usage in 1972. G.711 represents logarithmic pulse-code modulation (PCM) samples for signals of voice frequencies, sampled at the rate of 8000 samples/second (from wikipedia.org).

ITU-T G.723.1: A dual rate codec that compresses the speech signal to 6.3 and 5.3 kbps while maintaining the toll quality and relatively low delay (from ieeexplore).

ITU-T G.729. G.729: Is an audio data compression algorithm for voice that compresses voice audio in chunks of 10 milliseconds. Music or tones such as DTMF or fax tones cannot be transported reliably with this codec, and thus use G.711 or out-of-band methods to transport these signals. G.729 is mostly used in Voice over IP (VoIP) applications for its low bandwidth requirement. Standard G.729 operates at 8 kbit/s, but there are extensions, which provide also 6.4 kbit/s and 11.8 kbit/s rates for marginally worse and better speech quality respectively. Also very common is G.729a which is compatible with G.729, but requires less computation. This lower complexity is not free since speech quality is marginally worsened (from wikipedia.org).

ITU-T H.323: A protocol suite defined by ITU-T, is for voice transmission over internet (Voice over IP or VOIP). In addition to voice applications, H.323 provides mechanisms for video communication and data collaboration, in combination with the ITU-T T.120 series standards. H.323 is one of the major VOIP standards, just as Megaco and SIP. H.323 is an umbrella specification, because it includes a various other ITU standards (from cisco.com).

Jitter: Jitter is an unwanted variation of one or more signal characteristics in electronics and telecommunications. Jitter may be seen in characteristics such as the interval between successive pulses, or the amplitude, frequency, or phase of successive cycles. Jitter is a significant factor in the design of almost all communications links (from wikipedia.org).

Key Performance Indicator (KPI): Key Performance Indicators (KPI) are financial and non-financial metrics used to help an organization define and measure progress toward organizational goals. KPIs are used in Business Intelligence to assess the present state of the business and to prescribe a course of action. The act of monitoring KPIs in real-time is known as business activity monitoring. KPIs are frequently used to “value” difficult to measure activities such as the benefits of leadership development, engagement, service, and satisfaction (from wikipedia.org).

Mean Opinion Score (MOS): A common measure 8from 1 (bad) to 5 (excellent)) for subjective speech quality is the Mean Opinion Score (MOS) scale, defined in the ITU-T standard P.800. In a MOS test, the test persons listen to short speech samples, where every speech sample consists of two to five sentences. The total MOS score is then the mean of all individual results. Due to the absolute nature of the grading, this kind of test is also called an Absolute Category Rating (ACR) test. (from ericsson.com).

Public Switched Telephone Network (PSTN): Is the network of the world’s public circuit-switched telephone networks, in much the same way that the Internet is the network of the world’s public IP-based packet-switched networks. Originally a network of fixed-line analog telephone systems, the PSTN is now almost entirely digital, and now includes mobile as well as fixed telephones. The PSTN is largely governed by technical standards created by the ITU-T, and uses E.163/E.164 addresses (known more commonly as telephone numbers) for addressing (from wikipedia.org).

Quality of Service. Quality of Service (QoS): Refers to the capability of a network to provide better service to selected network traffic over various technologies, including Frame Relay, Asynchronous Transfer Mode (ATM), Ethernet and 802.1 networks, SONET, and IP-routed networks that may use any or all of these underlying technologies. The primary goal of QoS is to provide priority including dedicated bandwidth, controlled jitter and latency (required by some real-time and interactive traffic), and improved loss characteristics. Also important is making sure that providing priority for one or more flows does not make other flows fail. QoS technologies provide the elemental building blocks that will be used for future business applications in campus, WAN, and service provider networks (from www.cisco.com documentation).

Real-time Protocol (RTP): The Real-time Transport Protocol (or RTP) defines a standardized packet format for delivering audio and video over the Internet. It was developed by the Audio-Video Transport Working Group of the IETF and first published in 1996 as RFC 1889 which was made obsolete in 2003 by RFC 3550. Real time transport protocol can also be used in conjunction with RSVP protocol which enhances the field of multimedia applications (from wikipedia.org).

Session initiation Protocol (SIP): The Session Initiation Protocol (SIP) is an application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. It can be used to create two-party, multiparty, or multicast sessions that include Internet telephone calls, multimedia distribution, and multimedia conferences. (cit. RFC 3261). SIP is designed to be independent of the underlying transport layer; it can run on TCP, UDP, or SCTP. It was originally designed by Henning Schulzrinne (Columbia University) and Mark Handley (UCL) starting in 1996. The latest version of the specification is RFC 3261 from the IETF SIP Working Group. In November 2000, SIP was accepted as a 3GPP signaling protocol and permanent element of the IMS architecture (from wikipedia.org).

Talkspurt: A single unit of silence between two periods of speech.

User Datagram Protocol (UDP): User Datagram Protocol (UDP) is one of the core protocols of the Internet protocol suite. Using UDP, programs on networked computers can send short messages sometimes known as datagrams (using Datagram Sockets) to one another. UDP is sometimes called the Universal Datagram Protocol. It was designed by David P. Reed in 1980. UDP does not guarantee reliability or ordering in the way that TCP does (from wikipedia.org).

Chapter XLV
Open APIs and Protocols for Services and Applications in Telecoms

Nikolaos D. Tselikas
University of Peloponnese, Greece

Georgia M. Kapitsaki
National Technical University of Athens, Greece

Dimitrios G. Makris
University of Peloponnese, Greece

Iakovos S. Venieris
National Technical University of Athens, Greece

ABSTRACT

The role of open Application Programming Interfaces (APIs) and protocols for advanced service provisioning and the corresponding state of the art are the main subject of this chapter. Specifically, the role and the trade-offs in modern telecoms between open APIs and Protocols, that is, OSA/Parlay APIs, JAIN APIs and SIP, are discussed. A technical implementation analysis for each solution is presented, based on a call-related service, in order to set a common basis for the aforementioned technologies, since "voice" is a common denominator for a Fixed or Mobile Operator or an Internet Service Provider. A performance evaluation study regarding the implemented services is also presented and the chapter is summarized by interesting conclusions and related future trends.

1. INTRODUCTION

The demand for telecom services moving towards more and richer value added capabilities is an exceptional revenue opportunity for network operators. Traffic flowing through their networks will increase as more and more services for end-users become available. Network operators, however, are not able to deploy in their network services as fast as demand requires because core network interactions tend to be complex. Integrating new services directly into the network core and, most importantly, administering and maintaining those services after deployment, inevitably becomes a slow and cumbersome procedure. Rapid service creation and deployment is achieved only if service development, deployment and administration are distributed among external service providers. The most efficient way to achieve such a distribution is for network operators to expose parts of their network's core functions to outside providers.

Providing access to core network functions to outside providers requires compromises from both sides, in order to address the concerns raised by both parties. Network operators are concerned about security and stability, so they require control over interactions, while service providers are concerned with the effort required to develop new services and whether their investment in the developed software will be reusable by other operators. In the interests of both, interactions with the network should be simple to comprehend and implement. In addition they should require minimal development and integration effort and the result should be reusable, in order to protect the investment.

The proper way to address these concerns is by designing clear and simple interfaces between operators and providers. These interfaces define a simplified model of core network functions and are standardized to allow reusability and independence from underlying network architectures. Such interfaces can be implemented by choosing from a variety of communication middleware technologies; most notably distributed, object based systems such as RMI, CORBA, SOAP with RPC semantics or messaging frameworks like Java Message Service (JMS).

The primary concern using this approach has always been the performance impact produced by the additional middleware layers compared to the straightforward case of deploying services directly into the network's core. Therefore, a comparison of various middleware technologies supporting open interfaces is an important step before choosing the most appropriate communication middleware that will be used to implement an open interface for a specific value added service. This is the issue discussed in this book chapter; namely, to make a detailed survey, a in-depth investigation and an efficient performance comparison between several similar middleware systems based on open APIs and Protocols for Service Provisioning.

The rest of this chapter is organized as follows. The second subsection defines the role of open Application Programming Interfaces (APIs) and protocols for advanced service provision and the current state of the art in the field. Specifically, the role and the trade-offs in modern telecoms between open APIs and Protocols such as OSA/Parlay APIs, JAIN APIs, SIP and middleware technologies such as CORBA and RMI are addressed. This is followed by the third subsection attempts to set a common basis, in order to make possible a comparison between the different platforms and technologies. A call-related service and how this service can be implemented based on different open APIs and middleware technologies (i.e. OSA/Parlay API, JAIN API and SIP) is described, since "voice" is a common denominator for either Fixed or Mobile Operators or Internet Service Providers. The fourth subsection presents a performance analysis of the implemented services and the fifth and last subsection summarizes the preceding subsections,includes some unexpected conclusions and discusses related future trends.

2. SERVICE PROVISION IN TELECOMS

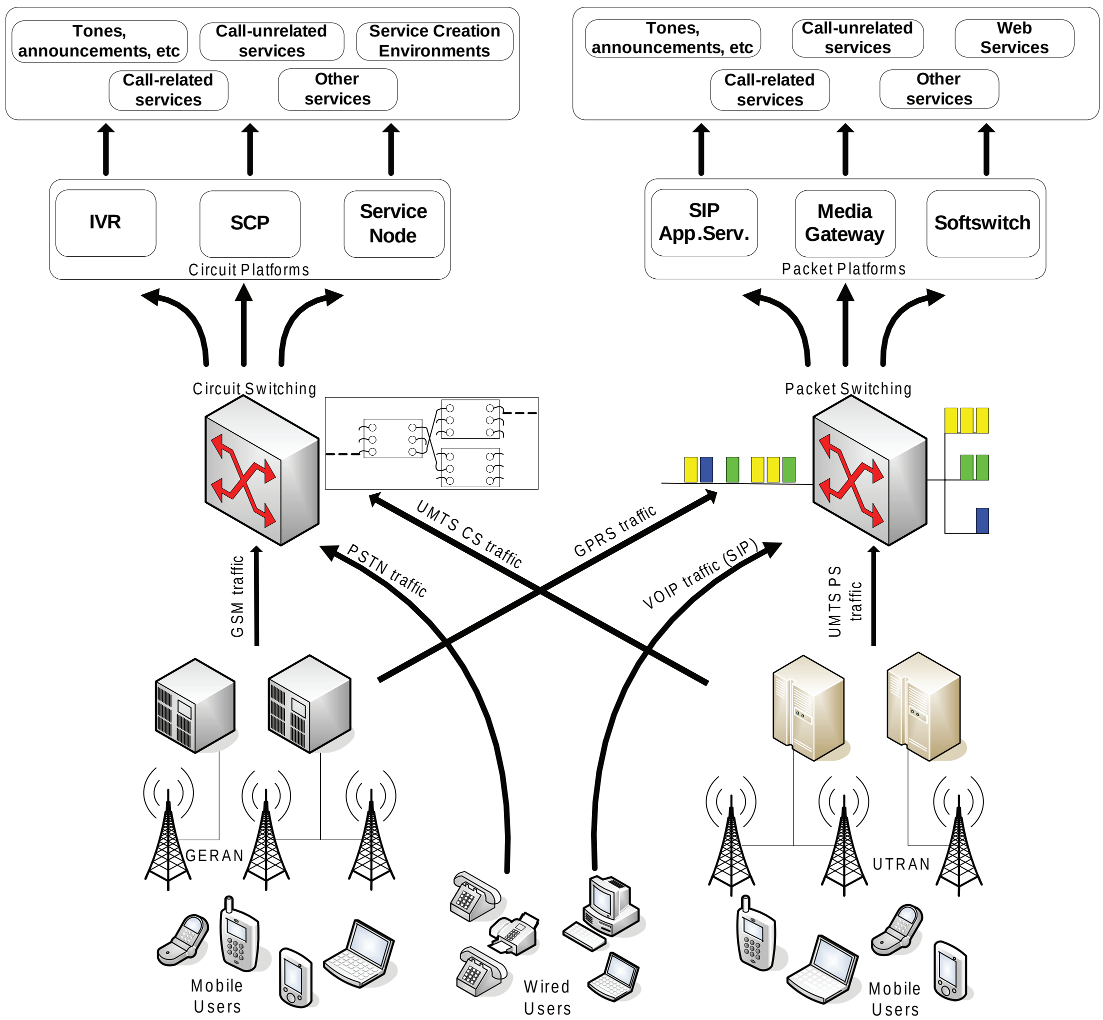
One of the main concerns of telecommunication networks has always been and still remains to propose and provide new services and applications. Multiple heterogeneous Access Networks – both wired and wireless – co-exist, allowing interactions between them. The ultimate goal of these networks is to provide fast and transparent access to modern services in the framework of the so-called NGN (Next Generation Network). Different networks provide different capabilities for their users and each present their own peculiarities (Cochennec, 2002). This kind of heterogeneity reflects the pros and cons that each network presents on its service provisioning plane (Danfeng, 2007).

The classification of telecommunication networks into circuit- and packet- switched introduces a major division that affects service provisioning (Figure 1). Different architectures, based on specific protocol implementations, have been adopted.

The IN (Intelligent Network) model for PSTN service provisioning, proposed new physical and functional entities that communicate with each other via INAP (Venieris, 1998). The corresponding architecture, CAMEL (Customised Application for Mobile network Enhanced Logic) used in the PLMN (Public Land Mobile Network) case for GSM and GPRS service provisioning is also built upon entities and protocols inspired by the IN model (MAP and CAP) (Cookson 2001). The idea of building a packet-switched VoIP network, based on the strict circuit-switched networks' rules, was brought into effect via the H.323 VoIP architecture, but with limited success because of these strict rules which are in conflict with the freedom that Internet Protocol represents (ITU-T Rec. H.323; Schulzrinne, 2000). The afore-mentioned specific and restricted solutions were both network and protocol oriented, thus they couldn't be applied to service provisioning over different kind of networks. Furthermore, as networks evolved, new players came onto the scene, that had to be taken into account for a wider perspective of service provisioning. Apart from the Network Operator, which used to own both the core Network and the Services, Independent Service Providers (ISPs) are playing an increasingly larger role in modern telecommunications by using existing network infrastructures to provide services under their own management. As a result, the final link of the telecommunication market's chain, i.e. the users, are theoretically able to select the services and applications they really need or desire among a wide range of services offered by different Service Providers and Network Operators.

In order to respond to this realization and satisfy users' demand for new and advanced services, a different approach has been followed in the last several years regarding service provisioning. This new approach is not another vertical one, but tends to be applied in as much a horizontal manner as possible. This horizontal approach attempts to cover and hide the underlying network peculiarities in the service plane by exposing only the actual necessary functionality for the creation and provision of new services. This is offered by the so-called Service Platforms, which operate over different network infrastructures. They are responsible for service deployment, manipulation, control and provisioning, while providing the basis for external service creation by Independent Service Developers and Providers. Service platforms achieve communication transparency by exploiting middleware technologies like CORBA, Java-RMI, Web Services technologies, etc (Venieris, 2000). Moreover, a standards-based approach in such platforms is imperative for providing Independent Service Providers with homogeneous access to the underlying network resources. These standards must be open, flexible and easily programmable for everyone; therefore, heavy and maladjusted protocols are avoided in Service Platform implementations. On the other hand, open APIs, such as OSA/Parlay and Java

Figure 1. Vertical approach to service provisioning



APIs for Integrated Networks (JAIN) or standard Internet-based Protocols like the Session Initiation Protocol (SIP) are preferred in many implementations, due to the previously mentioned advantages (Carvalho, 2007).

2.1 OSA/Parlay APIs

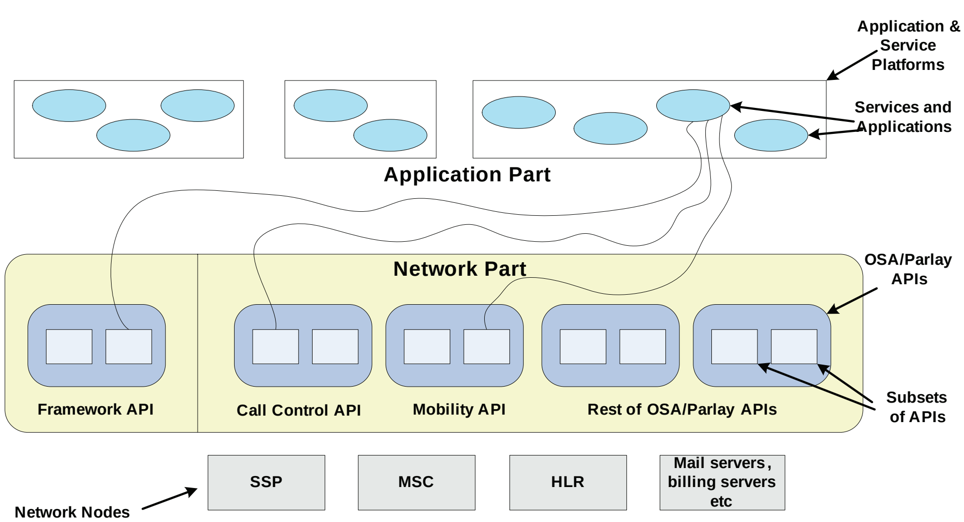
OSA/Parlay constitutes a prominent forum that acknowledges the importance of designing APIs for telecommunications capabilities (The Parlay Group). The Parlay APIs specifications have been developed and defined, allowing services and applications to access transparently the core network functionality. In the mobile world, the 3rd Generation Partnership Project (3GPP) adopted the Parlay specifications prior to specifying, defining and creating the Parlay-like APIs for the support of service development on top of mobile networks (3GPP; Rautela, 2002). This initiative is known as the OSA Interface, or just OSA (Open Service Access). After the release of Parlay's APIs version 3.0, Parlay and OSA merged in the same forum with the combined name "OSA/Parlay".

The first activities regarding OSA/Parlay took place in 1998 by a group of companies including BT, Microsoft, Nortel, Siemens and Ulticom known as the Parlay Group. Their goal was to devise an API that would provide 3rd parties with access to the network functions required for the creation and deployment of value added services. Over the years the OSA/Parlay Group has grown in popularity and currently has enlisted more than 10 full and more than 45 affiliate members.

OSA/Parlay (The Parlay Group) provides 3rd parties with programming interfaces that are open, language and platform independent, and include security provisions. Thus, OSA/Parlay APIs can be supported on top of various middleware technologies such as CORBA (Common Object Request Broker Application), DCOM, Java-RMI and the Simple Object Access Protocol (SOAP). OSA/Parlay defines two roles in the process of service provisioning via the corresponding parts of the OSA/Parlay APIs: One of these is the Network Operator, that implements the Network Part of OSA/Parlay APIs and the other is the Application Service Provider (ASP) that implements the Application Part of OSA/Parlay APIs. The Network Operator is typically perceived as deploying and operating the actual network infrastructure, e.g. switching and transmission equipment, access facilities, etc. The ASP provides numerous networks with value added services by gaining access to the Network Provider's resources via the OSA/Parlay APIs. The services provided by the ASP are executed on their application servers. The application servers become clients to one or more OSA/Parlay gateways which reside at the network edge offering Service Control Functionality (SCF) to the client via the OSA/Parlay API. A high-level OSA/Parlay architecture is depicted in Figure 2.

The special case of the Web Services-oriented OSA/Parlay specifications is implemented with Parlay X APIs. Parlay X started with its first version in March of 2003 as a subset of the Parlay APIs functionality providing the ability for application developers to have access to the Parlay gateways using Web Services technology. Only a few years later – i.e. in the current version v.3.0 – its functionality has been enhanced and currently Parlay X is more powerful than the traditional OSA/Parlay API, because of the Web Services revolution. One of the main advantages of Web Services technology is that it is both a language and platform independent technology. Both sides, the client and the server, can be developed completely independently using separate technologies, e.g. a Java client can talk to a C# server, etc. Service discovery is achieved by publishing a WSDL (Web Services Description Language)

Figure 2. A high-level OSA/Parlay architecture



file independently (W3C) and the users of the service can be directed to this WSDL file. It can be said that WSDL is the description of services in XML-format, similar to the corresponding Interface Definition Language (IDL) used in CORBA. The WSDL file describes how the service can be used. It gives details on the available operations, attributes and return types.

The OSA/Parlay APIs and Parlay X APIs are designed to enable the creation of telephony applications as well as to "telecom-enable" IT applications, without requiring developers to have an indepth understanding of telecommunications concepts. IT developers, who create and deploy applications outside the traditional telecommunications network space and business model, are viewed as crucial for creating dramatic market growth in the next generation of applications, services and networks. Thus, the OSA/Parlay APIs and Parlay X Web Services are intended to stimulate the development of next generation network applications by IT developers who are not necessarily experts in telephony or telecommunications (Lofthouse, 2004).

2.2 JAIN APIs

Java APIs for Integrated Networks (JAIN) is defined and specified by a large number of participating telecommunication firms, the so-called JAIN Community (Java Technology). The JAIN Community envisioned the creation of a number of open Java APIs that abstract the details of network functions and protocol implementations, in order to ease the development of portable applications. JAIN provides a Java-based framework to build and integrate services and solutions that span both packet- and circuit- switched networks. This was the major reason for the creation of the JAIN Protocol Experts Group (PEG) (Bhat, 2000). JAIN PEG focuses on developing Java APIs for protocols used in telephony, intelligent networks (INs), wireless networks, and the Internet. JAIN PEG is organized into two major divisions; the Signaling System No. 7 (SS7) subgroup and the Internet Protocol (IP) subgroup. The former focuses on developing Java APIs for SS7 technology mainly used in telephony, IN, and wireless networks, while the latter focuses on developing Java APIs for Internet technologies. Within each subgroup there are Edit Groups that focus on specific protocols.

Inspired by the OSA/Parlay forum, but strictly in the context of the Java language, the main objective of JAIN is to provide service portability, convergence, and secure access (for those services residing outside of the network) to such integrated networks, while the ultimate target of the JAIN Community is to create an open market for services across integrated networks using the already widely accepted Java technology (Tait, 2000).

2.3 Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP) is an application layer text-based protocol standardized by the Internet Engineering Task Force (IETF) in early 1999 (Rosenberg, 2002). It is used for session initiation, modification and termination among two or more terminals. Applications based on SIP focus on interactive multimedia sessions, such as Internet phone calls or multimedia conferences, but the protocol can also be used for instant messaging, event notification or managing other session types, such as distributed games. In setting up sessions, SIP acts as a signaling protocol, offering services similar to telephony signaling protocols such as Q.931 or ISUP, but in an Internet context (Schulzrinne, 2000). SIP also uses existing IETF protocols to support various applications (e.g. voice, call control). In combination with the Session Description Protocol (SDP) (Handley, 2006), SIP can describe the session characteristics, while, at the same time, signaling and media streams are separated. Due to its flexibility, many extensions have been proposed for enhancements to the supported functionalities.

SIP, in conjunction with the proposed extensions, supports many call control services, such as Call Forwarding, Call Transfer, Call Hold, Call Waiting, Call Identification, Conferencing and Third Party Call Control, as well as new Internet-based services like Click-to-Dial, capability exchange, distributed gaming, instant messaging and presence. Because of its increased simplicity in implementation – in contrast to H.323 – SIP is the dominant architecture for VoIP telephony (Papadakis, 2004).

2.4 Open APIs and Standard Protocols in Telecom Industry

Open APIs and standard protocols have already penetrated the Telecom Market. Vendors such as Ericsson, Siemens, Alcatel and Aepona have already incorporated OSA/Parlay APIs into their Service Platforms, i.e. Jambala, @dvantage, A-8601 and Causeway Parlay Gateway respectively (Ericsson Jambala Parlay SCS; Nokia Siemens Parlay @vantage; Alcatel 8601 Parlay/OSA Gateway; AePONA Causeway). Open Cloud and Sun have implemented JAIN in their corresponding Service Platforms; Rhino and jNETx, while the latter actually implements both JAIN and OSA/Parlay APIs in parallel (Java Technology ; SUN jNETx OSA Platform). On the other hand, SIP is the referential protocol in the IP Multimedia Subsystem (IMS), which is an architectural framework for delivering internet protocol (IP) multimedia to mobile users. IMS was originally designed by the wireless standards body 3rd Generation Partnership Project (3GPP), and is part of the vision for evolving mobile networks beyond GSM and GPRS, such as WLANs, 3G and fixed line. All the aforementioned vendors provide, in parallel, SIP-based service platforms. Siemens (Nokia Siemens Networks since 2007) has extended the Parlay @vantage Service Platform to IMS@dvantage supporting SIP too (Nokia Siemens).

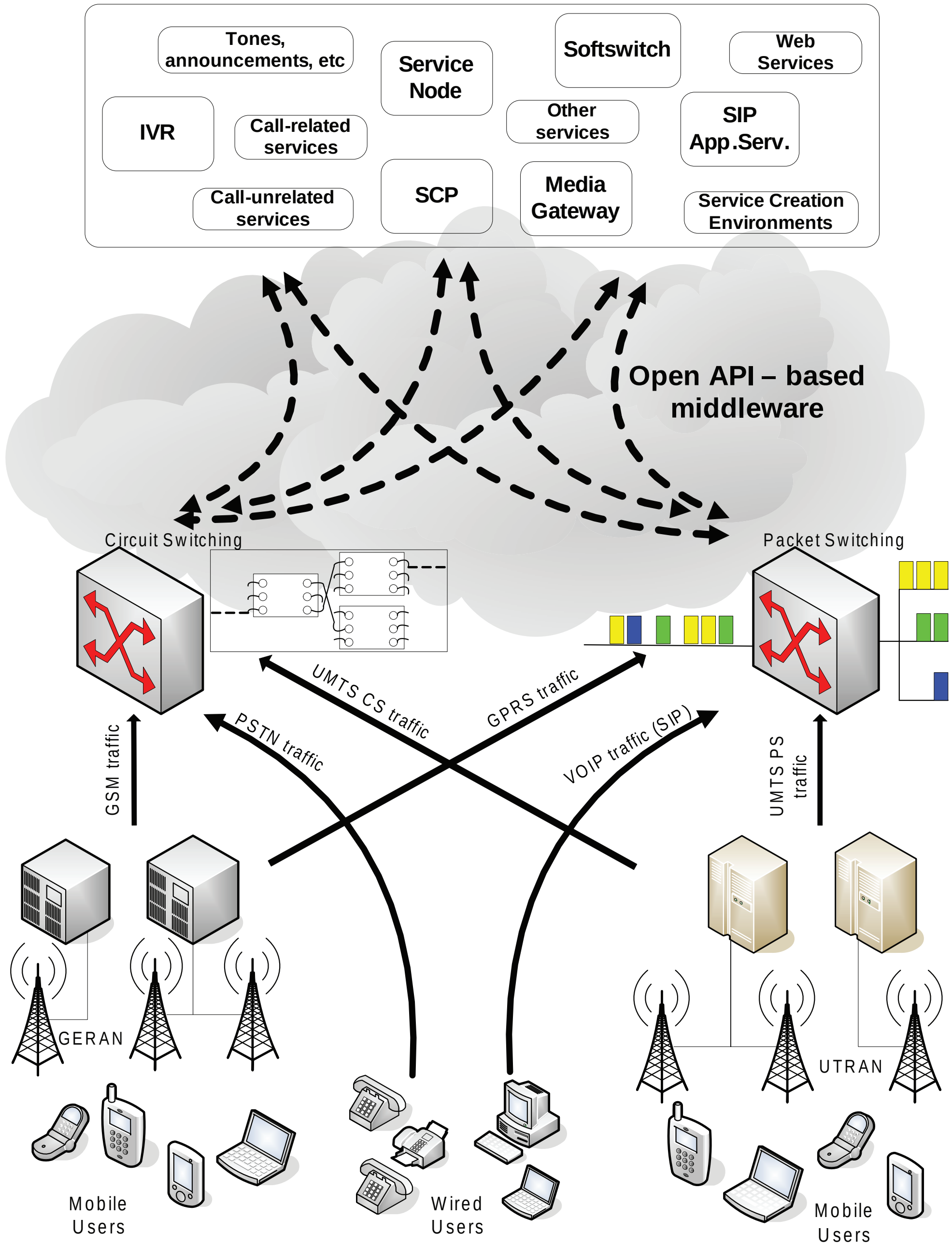
Such service platforms – or even middleware implementations developed from scratch – based on SIP or on open APIs like OSA/Parlay or JAIN, are used to create a cost-effective and highly-flexible IP-based infrastructure, in order to provide revenue generating services for the convergence of data, voice and mobile network technologies, as required in Next Generation Networks as depicted in Figure 3.

Regarding the Network Operators' and Service Providers' concern of making new and advanced services in an open, easier and more rapid manner, this kind of approach seems suitable (Moerdijk 2003). However, there are performance implications regarding middleware and the implemented open-API, when a new service is deployed for end-users. Subscribers to any service are always impatient. They demand services that respond as fast as possible with the minimum possible delay. But, advanced services may include several interactions between different network and application components, which are, sometimes, generated in real time introducing larger than desired delays.

3. IMPLEMENTING TELECOM SERVICES BASED ON OPEN APIS OR STANDARD PROTOCOLS

This section describes and analyzes a middleware implementation based on a subset of all three technologies discussed, i.e. OSA/Parlay APIs, JAIN APIs and SIP, in terms of call control functionality. The subsets of APIs involved are the Generic Call Control (GCC) API and the JAIN Call Control (JCC) API for OSA/Parlay and JAIN respectively. Performance evaluation results are also analyzed, in order to define whether (or not) an open API based middleware solution or SIP can consistently perform well. A simple call-related service is examined to determine the service response time by using the pre-mentioned solutions. The performance analysis is not based on the service processing time – which depends directly on the service logic and thus will vary

Figure 3. Service Provision in NGN



– but on an estimation of the raw performance of an open API based middleware or SIP stack implementation. This is actually the reason for not examining a more complex and advanced service. The limited observation of performance for such a simple call-control service renders insight into future expectations for real advanced services based on open API middleware solutions or SIP. Positive results can provide a boost in this service engineering trend.

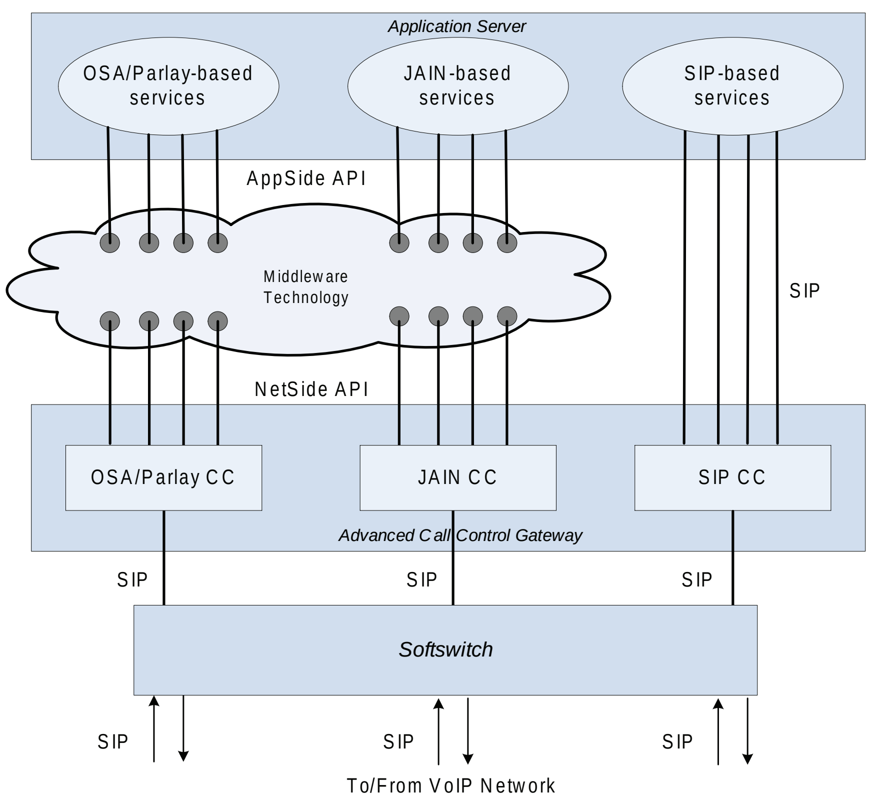
Based on the rationale described in the previous section, different implementations – realizing an advanced Call Control Gateway, based on different open APIs or SIP –will be presented, analyzed and evaluated. The advanced Call Control Gateway relies in parallel on all three types of Call Control, i.e. OSA/Parlay Generic Call Control, JAIN Call Control APIs and SIP Call Control. All three implementations have been tested regarding their performance over a commercial Vocaltec's Softswitch, for applying call related services in a pure SIP-based VoIP network (Tselikas, 2007). Similar Gateway implementations can be applied to both PSTN and PLMN access nodes, to certify the layer of abstraction provided by the usage of open APIs in fixed or mobile communication networks (EURESCOM 2001). Figure 4 presents a high level view of the service layer used during our experiments with the VoIP network.

The Call Control Gateway mediates between the VoIP Softswitch and the Application Server and communicates via SIP with the former and via open APIs (OSA/Parlay and JAIN respectively) or directly via SIP with the latter. For experimental purposes, services residing in the Application Server are triplicated supporting OSA/Parlay, JAIN and SIP. A detailed examination of each part of the Call Control Gateway implementation is following.

3.1 OSA/Parlay Call Control API Implementation

The first part of the functionality of the advanced Call Control Gateway adopts the OSA/Parlay Generic Call Control Service and specifically the Generic Call Control API. The selection of the Generic Call Control API (version 4.1) depends on the requirements of the services that are going to be offered. Since the target is not to provide a very complex service, but to evaluate the performance of the open API middleware implementation, the

Figure 4. A high-level view of the experiment implementation



implementation of a simple service is more than enough for the experimental part of the study. Offering more advanced and complicated services such as Multiparty or Conference Call Control API could be implemented instead of Generic Call Control API.

OSA/Parlay APIs are divided in two parts, called "sides", namely the Network Side API and the Application Side API, respectively. The former is implemented in the Gateway, while the latter is implemented in the Application Server and is offered by the actual service provider. The Call Control model of OSA/Parlay is based on the traditional IN call model (Hellenthal, 2001). The requirement for the implementation of the Generic Call Control API is to provide two interfaces on the Network Side, namely the `IpCallControlManager` and the `IpCall`.

From an implementation point of view, the two Network Side interfaces can be considered as two CORBA or RMI servers exposing their methods to the Application Part, since OSA/Parlay provides the corresponding interfaces' definitions in "IDL" (Interface Definition Language), hence DOT (Distributed Object Technologies) like CORBA or RMI can be easily adopted (Venieris, 2000). Actually, the two Network Side objects (`IpCallControlManager` and `IpCall`) constitute the OSA/Parlay part of the Call Control Gateway, which is responsible for providing an OSA/Parlay-oriented view of network resources and recognizing the respective Application Side OSA/Parlay objects as seen by the signatures of their methods. The latter are the `IpAppCallControlManager` and `IpAppCall`, which expose the callback methods to the Network Side objects' methods. Briefly, the most common of them are: the `IpAppCallControlManager::callEventNotify()` method, which notifies the application of the arrival of a call related event, the `IpAppCall::routeRes()` method, indicating that the routing request to the destination party was successful, and finally the `IpAppCall::routeErr()` method, which indicates that the routing request to the destination party was unsuccessful, as well as the corresponding reason for the failure. Figure 5 provides a picture of the OSA/Parlay part of the advanced Call Control Gateway in terms of a UML class diagram.

Figure 6 depicts the sequence diagram of the

Figure 5. UML class diagram of OSA/Parlay GCC API implementation

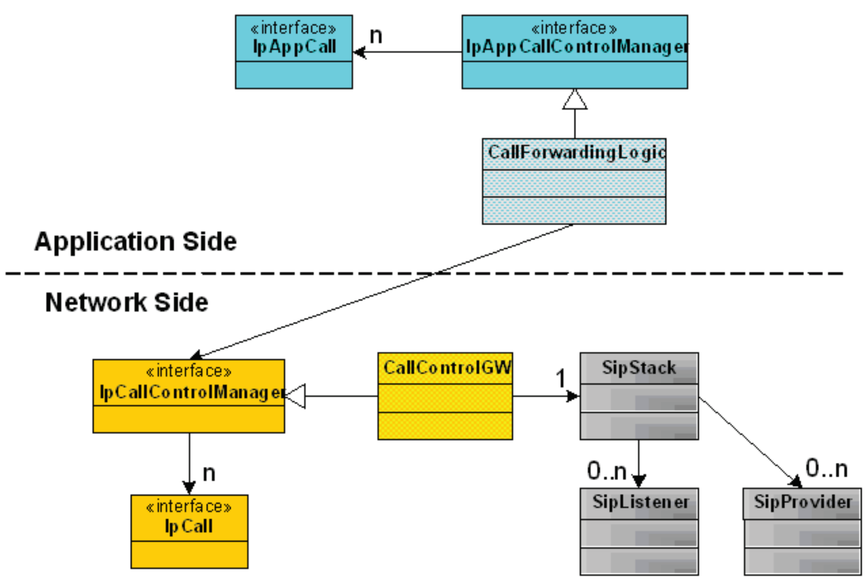
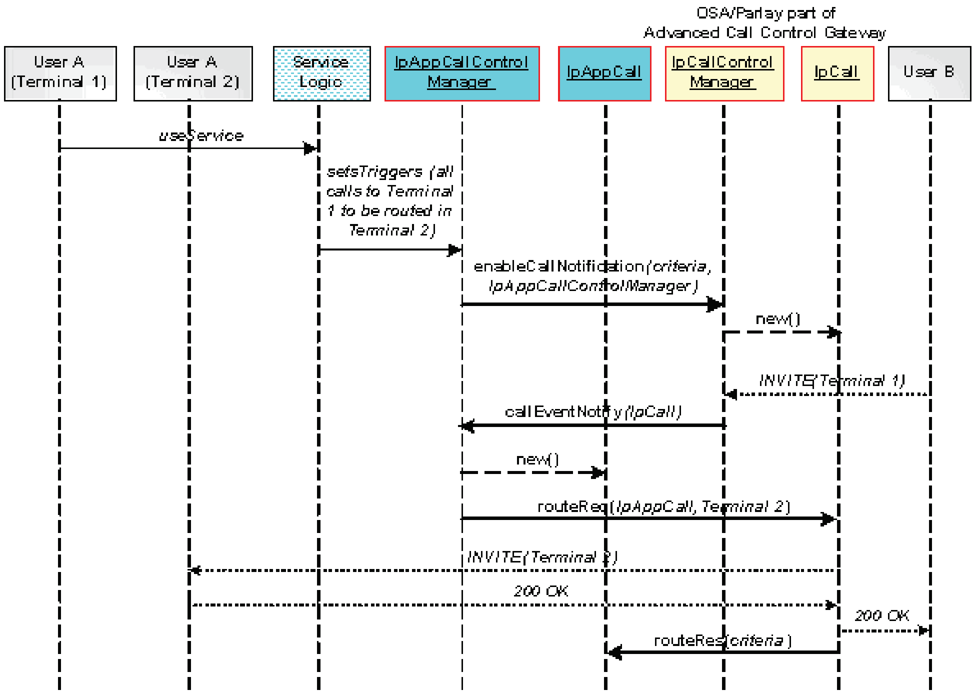


Figure 6. OSA/Parlay-based execution of "Call Forwarding" service (sequence diagram)



implemented "Call Forwarding Service", which clarifies the interaction between Generic Call Control Interface's objects.

By using the "Call Forwarding Service", a user is able to specify in which terminal he/she desires his/her calls to be routed (VoIP or, PSTN or PLMN). Both the Network Side and Application Side objects are implemented as persistent CORBA servers. The same implementation was established by implementing the afore-mentioned objects as RMI servers too, in order to compare the performance of the OSA/Parlay API when using different middleware technologies (i.e. CORBA and RMI). Since User A has already used the service, a Triggering Detection Point (TDP) has been set by the Application Side of OSA/Parlay via the method `enableCallNotification(criteria,IpAppCallControlManager)`. Then, the `IpCallControlManager` object generates a new `IpCall` object. In addition, the `IpCallControlManager` returns to the `IpAppCallControlManager` the reference to this new `IpCall` object via the `callEventNotify(IpCall)` method, when the relevant conditions or criteria have been detected and satisfied (User B calls the Terminal 1 of User A). At this point in time, service provisioning begins, while the `IpAppCallControlManager` generates a new `IpAppCall` object, and passes the criteria and the reference of `IpAppCall` to `IpCall` via `routeReq(criteria,IpAppCall)` method, for call routing to another destination that the callee has already set. The destination can be either another telephone device (another terminal number) or a mail account. Since the Event Detection Point (EDP) has been set, the `IpCall` object finally returns the result to `IpAppCall` via the `routeRes(criteria)` method, in order to complete the interaction for the service provisioning. In the first case (routing to another telephone device) the call is routed directly to the alternative number where the caller can be reached. In the second case (routing via mail, which has not been implemented for the case study), the call is routed to an intelligent peripheral, in which the caller can leave a voice message that will be routed to the callee's mail account either as an attachment file, or as a text mail if the intelligent peripheral supports speech-to-text conversion. The routing to an e-mail address, presupposes that the Messaging Service of Parlay (Generic Messaging API) is implemented between the mail server and the intelligent peripheral. Thus, the intelligent peripheral uses the `IpMailBoxFolder::putMessage()` method of the Generic Messaging API, in order to post the mail message to the callee's mail folder.

The implementation of OSA/Parlay Generic Call Control Service API was integrated with VocalTec's VoIP Softswitch, which is able to manipulate the SIP-stack's messages. In order to transform the SIP-stack messages to the OSA/Parlay Generic Call Control API, we mapped the OSA/Parlay prescribed events to the corresponding SIP messages. Table I depicts the most important of the previously mentioned relationship between OSA/Parlay Generic Call Control API events and SIP messages.

With such a mapping, every OSA/Parlay based service can set triggering criteria for the Call Control Gateway. The OSA/Parlay part of the Call Control Gateway is then responsible for mapping the received OSA/Parlay events and propagating the corresponding SIP messages to the Softswitch. Consequently, the service developer only has to follow one simple rule: "plain usage of OSA/Parlay API's" to create new services. Thus, he/she is not obliged to be familiar with ad hoc service tools.

3.2JAIN Call Control API Implementation

The second part of the advanced Call Control Gateway functionality is based on JAIN and particularly on the Java Call Control (JCC) API. JCC is functionally very similar to the corresponding OSA/Parlay Call Control one. JCC itself is based on the language-neutral OSA/Parlay v.2.1 Enhanced Call Control Service (ECCS) specification which is an effort to harmonize the JAIN and OSA/Parlay call control Protocols. However, for the Java Community, JCC is the official Java instantiation of the OSA/Parlay call control (Java Call Control (JCC) API, 2001). Consequently, the Java Call Control model of JAIN is also following a priori the traditional IN call model, by supporting analogous state machines for the different kind of objects participating in the API.

A typical, but not essential, difference between JCC v.1.0 and the OSA/Parlay Call Control is the terminology of the participating objects. On the Network Side the OSA/Parlay `IpCallControlManager` is now called `JccProvider`, while the `IpCall` object is represented by two objects, `JccCall` and `JccConnection` respectively. `JccCo-`

Table 1. OSA/Parlay to SIP mapping

OSA/Parlay Call Control Event	SIP message
`P_EVENT_GCCS_ADDRESS_ANALYSED_EVENT`	`INVITE`
`P_EVENT_GCCS_ADDRESS_COLLECTED_EVENT`	`100 (Trying)`
`P_EVENT_GCCS_OFFHOOK_EVENT`	`Not mapped in SIP stack`
`P_EVENT_GCCS_CALLED_PARTY_BUSY`	`486 (Busy)`
`P_EVENT_GCCS_CALLED_PARTY_UNREACHABLE`	`480 (Temp. Unavailable)`
`P_EVENT_GCCS_NO_ANSWER_FROM_CALLED_PARTY`	`180 (Ringing)`
`P_EVENT_GCCS_ROUTE_SELECT_FAILURE`	`4xx (but not 486 or 480)`
`P_EVENT_GCCS_ANSWER_FROM_CALL_PARTY`	`200 (OK)`
`P_EVENT_GCCS_DISCONNECT`	`BYE`

nnection can be considered as the IpCallLeg object of OSA/Parlay Multiparty Call Control API, where a set of "legs" actually represents a call. On the other hand, on the Application Side of the OSA/Parlay IpAppCallControlManager is now called JccProviderListener, and – as on the Network Side, the IpAppCall object is divided in JccCallListener and JccConnectionListener objects respectively. The JCC objects are considered as RMI servers, since both JCC and RMI are Java oriented. Thus, RMI technology is preferred for such an implementation rather than another middleware technology (e.g. CORBA). The UML class diagram in Figure 7 depicts the corresponding JAIN part of the advanced Call Control Gateway.

Referring back to the "Call Forwarding Service" described in the previous subsection, the respective sequence diagram describing the service execution according to the JAIN JCC API is presented in Figure 8.

Both Network and Application Side objects are implemented as persistent RMI servers. Since User A has already used the service, a Triggering Detection Point (TDP) has been set for the Network Side by the Application Side of JCC via the method addConnectionListener (JccConnectionListener,criteria). Then, the JccProvider object generates a new JccCall and a JccConnection object. When the criteria have been detected and satisfied (User B calls Terminal 1 of User A) the JccConnection returns to the JccConnectionListener the reference of the appropriate JccCall object via the connectionAlerting(JccCall,event) method. This is the point in time when service provisioning starts up. The JccProviderListener, has already generated a JccCallListener and a JccConnentionListener object, and passes the criteria and the reference of JccCallListener to JccCall via routeCall(JccCallListenerm, criteria) method, for call routing to another destination that the callee has already set (i.e. Terminal 2). Then the JccCall generates a new JccConnection object (i.e. JccConnec-

Figure 7. UML class diagram of JAIN JCC API implementation

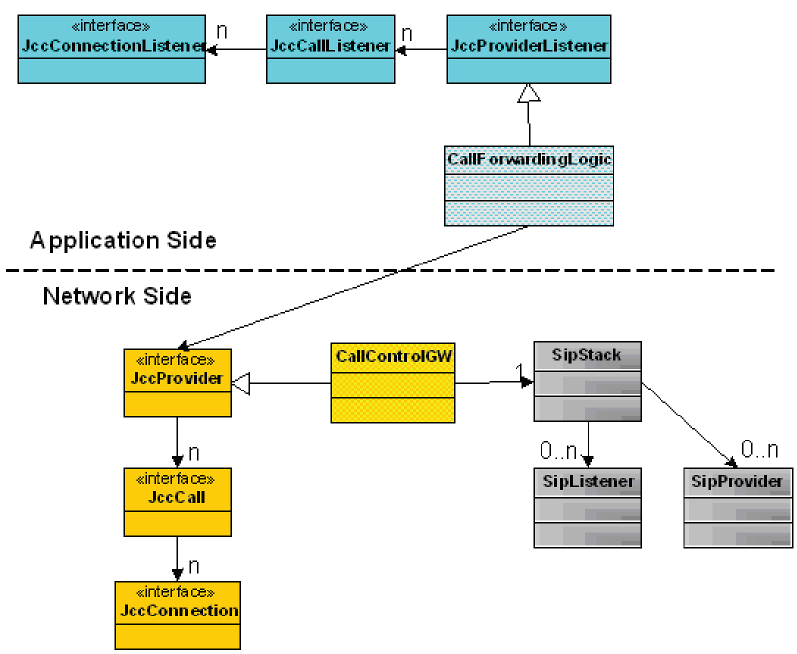
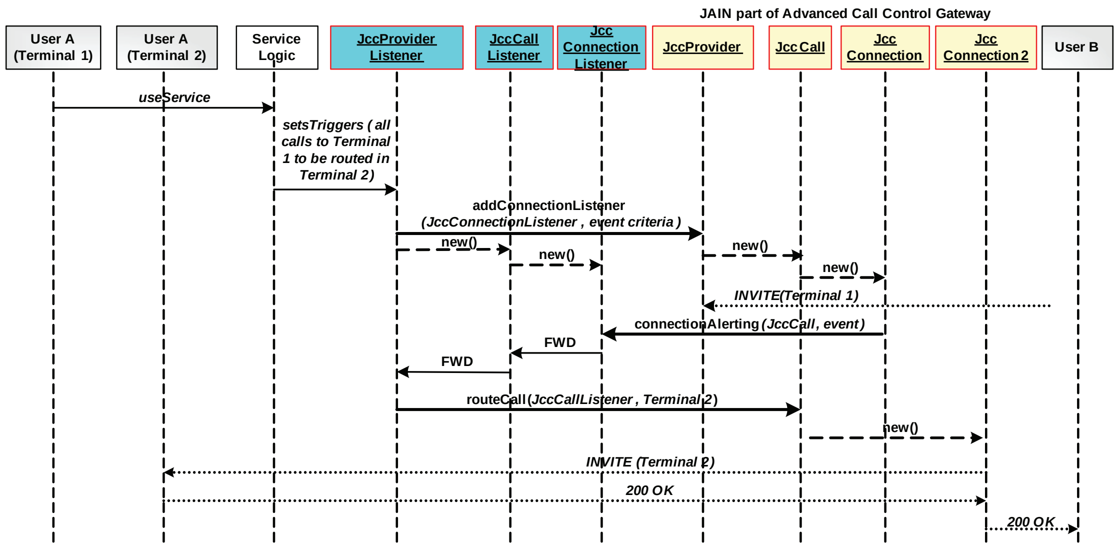


Figure 8. JAIN JCC-based execution of "Call Forwarding" service (sequence diagram)



`tion2` in Figure 8), representing the new call leg towards Terminal 2, which is responsible for forwarding the actual `INVITE` message to the new destination, in order to accomplish the requested service logic.

A short and quick comparison between Figure 6 and Figure 8 makes obvious the pre-mentioned correspondence between the OSA/Parlay GCC and JAIN JCC objects. In order to transform the SIP-stack messages to the JAIN JCC API, JCC events are mapped to the corresponding SIP messages. The most important of them, as well as their correlation with the SIP messages are presented in Table 2.

With such a mapping, the JCC-oriented services set the triggering criteria to the Call Control Gateway. The JAIN part of the advanced Call Control Gateway is then responsible for mapping the received JCC events and propagating the corresponding SIP messages to the Softswitch. Thus, a layer of abstraction is achieved. The one and only restriction is a result of the nature of this API and this is the Java language itself. But on the other hand, the Java language is the most preferable object oriented network programming language used by the majority of the developers.

3.3 SIP Call Control Implementation

The third type of advanced Call Control Gateway functionality presented here, describes a pure SIP implementation based on the JAIN SIP API (JSR 32, 2006), a full implementation of the SIP standard (J. Rosenberg, 2002). The essential difference between pure SIP and the other described approaches is that the interface between the call

Table 2. JAIN JCC to SIP mapping

JAIN JCC Event	SIP message
ADDRESS_ANALYSE	INVITE
ADDRESS_COLLECT	100 (Trying)
ALERTING	180 (Ringing)
SUSPENDED	480 (Temp. Unavailable)
FAILED	4xx (but not 486 or 480)
CONNECTED	200 (OK)
DISCONNECTED	BYE

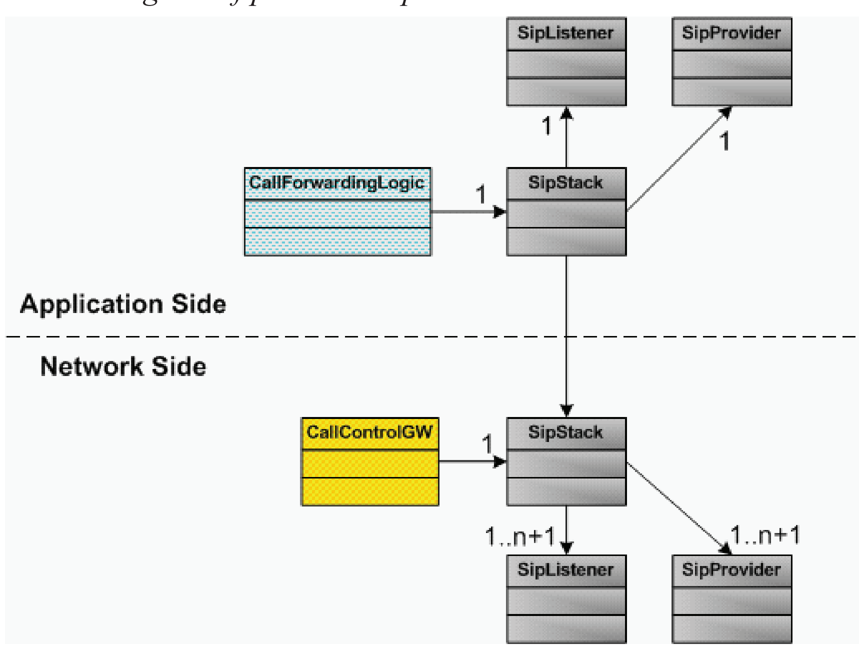
control server and the application server is based on the SIP protocol and, thus, the SIP signaling messages do not need to be translated to intermediate signaling messages.

On the Network Side, the Call Control Gateway CallControlGW instantiates a SIP stack, `SipStack`. The `SipStack` is responsible not only for the communication with the, n, SIP user agents, but also for the communication with the Application Side. Since the JAIN SIP architecture utilizes the Listener/Provider event model, n+1 `SipListener` and n+1 `SipProvider` objects are instantiated on the Network Side. On the Application Side, the `CallForwardingLogic` comprises the logic of the call forwarding service. The application-side, `SipStack,` is responsible only for the communication with the Network Side. Thus, one `SipListener` and one `SipProvider` object are instantiated. The UML class diagram in Figure 9 illustrates the pure SIP part of the advanced Call Control Gateway.

The "Call Forwarding Service" sequence diagram describing the service execution in the pure SIP case is presented in Figure 10. As soon as User A uses the service, the Service Logic sets a trigger for the Network Side, in order to redirect all SIP messages targeted for Terminal 1 to the Call Forwarding Logic. At this point in time, service provisioning starts up. SIP `INVITE` messages satisfying the call forwarding criteria are forwarded to Service Logic. The Service Logic is responsible for changing the appropriate header of the `INVITE` message (i.e. the first-line of the SIP request message and transmit it back to the Network Side) (Johnston, 2003). The `INVITE` message is not directly transmitted by the application-side SIP stack, since the call forwarding service may also have been triggered for Terminal 2.

Finally, it should be noted that although the testbed has been implemented in Java, the pure SIP approach is not limited by programming languages or the characteristics of the target machine. The Java language has been selected only for ease of implementation and in order to have a common basis regarding the performance evaluation (comparison) of the three different approaches.

Figure 9. UML class diagram of pure SIP implementation



4. PERFORMANCE ANALYSIS

This section describes a performance evaluation for the different service implementations. An indicative criterion useful for a direct comparison among all the three previously discussed implementations is the mean value of the measured total time for a whole life cycle of a service request ($\overline{t_{TOTAL}}$). The whole life cycle of a service request is defined as the required time period starting with the arrival of a SIP request in the Softswitch and ending with the successful response (by the Service Logic) arriving in the Softswitch respectively. $\overline{t_{TOTAL}}$ is estimated for the following cases, as described in the previous subsections, i.e. OSA/Parlay API using CORBA, OSA/Parlay API using RMI, JAIN API using RMI and plain SIP implementations respectively. The model used for the generation and arrival of service requests is based on the Poisson

distribution $\left(\frac{\lambda^k e^{-\lambda}}{k!}\right)$.

Five reasonable classes of different request rates have been taken into account, representing the λ-value of the Poisson rate. These are 20, 40, 60, 80 and 100 service requests per minute respectively. It should be noted that each experiment lasted about three hours, in order to derive reliable results, using the Poisson distribution.

The cluster columns bar chart in Figure 11 presents the mean total service time for each one of the five different rate experiments for all four cases (usage of OSA/Parlay Call Control API over CORBA and RMI respectively, usage of JAIN JCC API over RMI and usage of plain SIP). For lower traffic loads (i.e. 20, 40 and 60 requests per minute) JAIN JCC API over RMI show the most efficient implementation concerning $\overline{t_{TOTAL}}$, while both SIP and CORBA implementations of OSA/Parlay GCC API are less efficient and the corresponding RMI-based comes is least efficient. For higher traffic loads (i.e. 80 and 100 requests per minute) the SIP implementation is the most efficient, followed by JAIN and OSA/Parlay over CORBA and RMI respectievly.

Figure 10. Pure SIP-based execution of "Call Forwarding" service (sequence diagram)

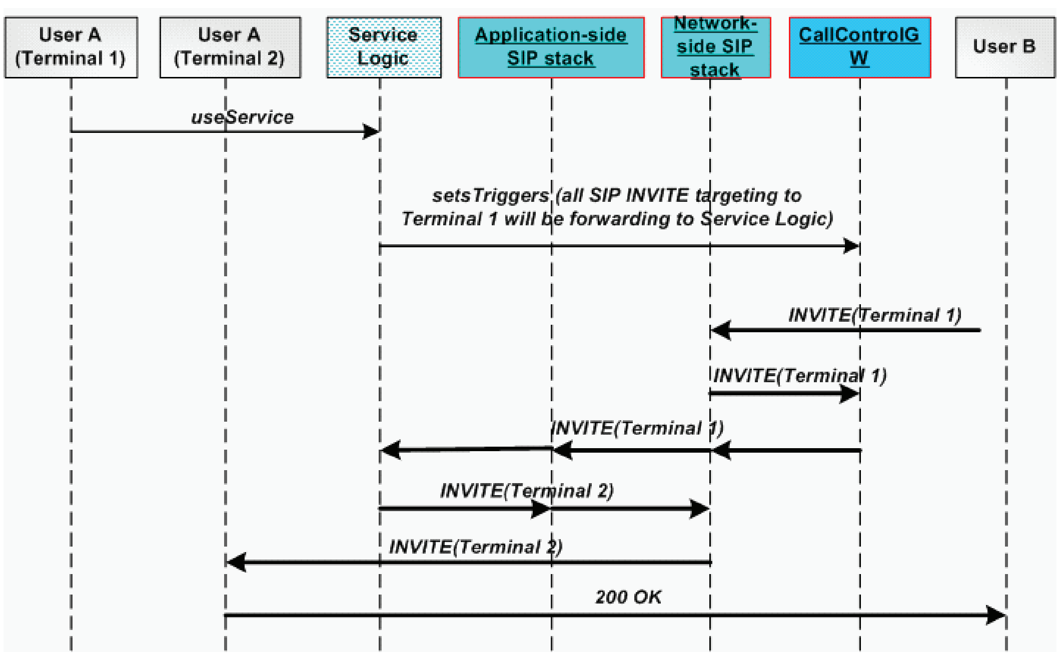
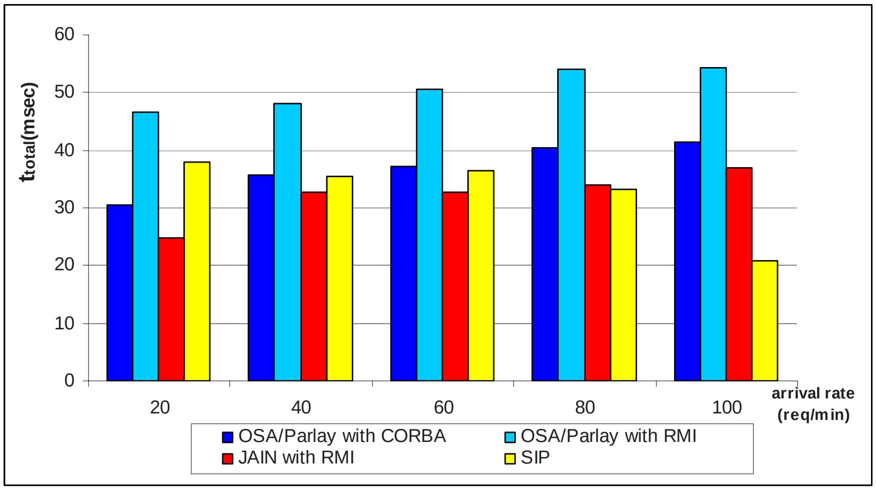


Figure 11. Mean Total Serving Time



An important observation is that for all three middleware based implementations, i.e. the implementations based on OSA/Parlay and JAIN API, $\overline{t_{TOTAL}}$ is increased when the arrival request rate is increased. Exactly the opposite happens in the SIP case. How can this paradox be explained? The communication mechanism in all four cases has the same basis. For the communication establishment over the RMI communication link, actually JRMP is used, which is TCP/IP based. This communication bus is used in two out of four experimental cases (OSA/Parlay GCC API over RMI and JAIN JCC API over RMI). In the third one (OSA/Parlay GCC API over CORBA), the CORBA bus of IIOP is used, which is also TCP/IP based. Finally, in the SIP case TCP sockets are also used for the communication. So the answer is not hidden in the communication mechanism, but probably in the time required in the three middleware-based cases to construct and de-construct the SIP `INVITE` messages to OSA/Parlay and JAIN methods and vice versa. The JAIN JCC over RMI case needs less time to construct the new structure of the JCC API method against the corresponding case of OSA/Parlay GCC API over RMI one. This can be explained, because OSA/Parlay APIs' methods contain more complicated data types and structures than the ones JAIN JCC API uses. On the other hand, JAIN JCC API is actually a Java-oriented implementation of OSA/Parlay GCC API, thus it is expected to be designed for RMI usage (RMI is a Java product too), while OSA/Parlay does not indicate a particular underlying middleware technology. In the SIP case, no such construction or de-construction is needed. Furthermore, RMI and SIP show better performance because a more efficient socket manipulation policy is used versus CORBA. A new RMI or SIP connection can either open a new socket or reuse an already opened one, which is currently idle (Stefano 2000). An already used RMI-socket stays alive for a few seconds. If – during this time – a new RMI connection is required, the already alive socket can be reused, saving time and resources, instead of opening a new one. Moreover, the possibility of having a lot of sockets alive is higher, when a lot of connections have been already established. This is probably another reason resulting in the better performance of JAIN and SIP. On the other hand, in the CORBA case, the underlying socket is destroyed after the usage, thus every new request arrival requires the generation of a new socket. Consequently, more time is spent for higher traffic loads.

The summarizing Table (Table III) gathers the main results of all four different approaches/implementations measured in the testbed used for the performance comparison of OSA/Parlay API, JAIN API and SIP protocol. The testbed used for the experiments consisted of two workstations, which are hosted by a Local Area Network (LAN) offering IP transport facilities. The first workstation – based on Windows XP Pro with a Pentium-4 1.8 GHz processor and 1 GB RAM – hosts the Call Control Gateway (i.e. Network Part of the OSA/Parlay GCC API, the JAIN JCC API and SIP implementations), as well as the SIP Proxy Agent, while the second one – based on Windows XP Pro with a Pentium-4 2.6 GHz processor and 1 GB RAM – hosts the SIP User Agents and the call control classes of the Application Part of the OSA/Parlay and JAIN implementations (i.e. the Application Part of OSA/Parlay GCC API and the "Listeners" of JAIN JCC API implementations respectively).

The full testbed software has been developed on pure Java (SDK and JRE 1.4.2). The SIP stack sed for the implementation of the SIP Proxy and User Agents was the JAIN-SIP 1.1 reference implementation. In the case where CORBA was used as the middleware communication platform between the OSA/Parlay network and application parts, the OSA/Parlay interfaces and data type classes were developed and deployed by the IDL compiler of the standard Java 1.4.2 SDK, and the ORB and Naming Service provided by the latter. In the case of the RMI middleware solution the OSA/Parlay and JAIN interfaces and data type classes were implemented independently and the RMI naming service (i.e. the RMIRegistry) provided by the standard Java runtime environment was used.

For the emulation of the Softswitch workload, SIP User Agents were used. They produce service requests, by generating SIP INVITE messages based on the Poisson distribution, as already mentioned. For each invitation message a unique id was used in order to distinguish the different service requests be able to measure the relative mean total times.

At this point one may notice that the traffic load used during the experiments does not correspond to the actual traffic load of a Service Provider. The answer is that the traffic load used is restricted by the available equipment, which provides limited network resources, since the

Table 3: Summarizing table

API / Protocol	OSA/Parlay API		JAIN API	SIP
Underlying Middleware Technology	CORBA	RMI	RMI	-
$\overline{t_{TOTAL}}$ low traffic load (i.e. 20 req/min)	30,61 msec	46,52 msec	24,69 msec	37,88 msec
$\overline{t_{TOTAL}}$ medium traffic load (i.e. 60 req/min)	37,31 msec	50,62 msec	32,78 msec	36,57 msec
$\overline{t_{TOTAL}}$ high traffic load (i.e. 100 req/min)	41,40 msec	54,32 msec	36,97 msec	20,85 msec
Re-usage of alive sockets	No	Yes	Yes	Yes
Technology dependency	Independent	Java-oriented	Java-oriented	Independent

testbed is based on plain PCs acting as servers and softswitches. In any case, these measurements give a clear indication for the potential deployment of the afore-mentioned approaches in commercial solutions (i.e. dedicated commercial servers and softswitches serving much higher traffic load), as well as a performance indication for other type of telecom-services (e.g. Location Based Services) based on similar APIs and underlying middleware technologies (e.g. OSA/Parlay Mobility API using CORBA or RMI). For example, in Location Based Services case using OSA/Parlay Mobility API, we expect higher performance using CORBA than using RMI, since OSA/Parlay Mobility APIs' and Call Control APIs' methods have approximately the same complexity and structure, thus we expect analogous treatment by the underlying middleware technology.

5. CONCLUSIONS

This chapter described the implementation of telecom services based on open interfaces and standard protocols exposing call control network functions. The current trend in service provisioning in telecoms was presented. An OSA/Parlay implementation with RMI and CORBA, a JAIN implementation using RMI as well as a pure SIP implementation were described, all with technical details. A performance evaluation study based on mean values of time measurements was also presented.

The most important conclusion one can draw from the performance analysis is that the performance impact imposed by the use of a middleware layer implementing open interfaces between the network and services is an acceptable overhead compared to the gains of exposing network functionality through Open APIs. The use of standardized interfaces to interact with the network is a significant factor contributing transparency, modularity, reusability and a clear distribution of administrative responsibilities to the service development process. This provides significant gains for all those involved; network operators increase the traffic and utilization of their networks through the use of novel services developed by 3rd parties; the simplified development and deployment process for services – transparent to the network – spurs activity in the service provider world and – as a consequence – users enjoy a large variety of value added services.

Extending the work that has been described in this book chapter, more advanced services could be developed via a more enriched service gateway. Examples might be the implementation of the User Interaction OSA/Parlay API or the OSA/Parlay Terminal Capabilities Service Features that enhance current functionality of VoIP services. The result would be a more complex gateway resembling an even more real-life deployment of advanced services supporting personalization of terminal capabilities and user location (Pailer 2003). Other middleware technologies can also be considered, like Parlay X APIs and the Web Services paradigm. A performance evaluation on gateways of such complexity would be a significant contributing factor to an insightful selection process of appropriate middleware technology supporting advanced value added services in telecoms. The goal is always the full exploitation of network capabilities by external providers, in order to eliminate its shortcomings. The solution offered to this problem by open APIs or standard protocols is the homogeneous and transparent service provision to any underlying network infrastructure.

REFERENCES

3GPP, at www.3gpp.org

AePONA Causeway, at www.aepona.com

Alcatel 8601 Parlay/OSA Gateway, at: www.alcatel.com

Bhat, R. R. & Gupta, R. (2000). JAIN Protocol APIs. *IEEE Communications Magazine, 38*(1), 100-107.

Carvalho, R. P., & Alberti, A. M. (2007). Java technologies for NGN service creation: Discussion and architecture to improve SIP addresses discovery. *Internet and Multimedia Systems and Applications Conference*, (pp. 56 – 62).

Cochennec, J. Y. (2002). Activities on next-generation networks under Global Information Infrastructure in ITU-T. *IEEE Communications Magazine*, *40*(7), 98-101.

Cookson, M. D., & Smith, D. G. (2001). 3G Service Control. *BT Technology Journal*, *19*(1), 67-79.

Danfeng, Y., & Fangchun, Y. (2007). A Universal Service Layer Management Model in NGN. *IEEE International Conference on Advanced Communication Technology*, 3, 1721-1725.

Ericsson Jambala Parlay SCS, at: www.ericsson.com

EURESCOM project P1110 (2001). *Open Service Access: Advantages and opportunities in service provisioning on 3G Mobile Networks.* Document Proposal for Enhancements to the OSA/Parlay specifications.

Handley, M., Jacobson, V., Perkins, C. (2006). *SDP: Session Description Protocol*. Request for Comments (Standards Track), 4566, Internet Engineering Task Force.

ITU-T Rec. H.323. (1998). *Packet-Based Multimedia Communications Systems.* Geneva.

Hellenthal, J. W., Panken, F. J. M., & Wegdam, M. (2001). Validation of the Parlay API through prototyping. *IEEE Intelligent Network Workshop.*

Java Call Control (JCC) API (2001). Version 1.0, *Overview of the API.* Java Technology. The JAIN initiative, at http://java.sun.com/products/jain/

Johnston, A., Donovan, S., Sparks, R., Cunningham, C., & Summers, K. (2003). *Session Initiation Protocol (SIP) Basic Call Flow Examples*. Request for Comments (Best Current Practice), 3665, Internet Engineering Task Force.

JSR 32 (2006). *JAIN SIP API Specification*. http://jcp.org/en/jsr/detail?id=32

Lofthouse, H., Yates, M. J., & Stretch, R. (2004). Parlay X Web Services. *BT Technology Journal*, 22(1), 81-86.

Moerdijk, A., & Klostermann, L. (2003). Opening the networks with OSA/Parlay APIs: Standards and aspects behind the APIs. *IEEE Network Magazine*, 17(3), 58-64.

Nokia Siemens Parlay @vantage, at http://www.nokiasiemensnetworks.com

Pailer, R., Stadler, J., & Miladinovic, I. (2003). Using PARLAY APIs Over a SIP System in a Distributed Service Platform for Carrier Grade Multimedia Services. *Wireless Networks,* 9(4), 353-363.

Papadakis, A. E., Chaniotakis, E .S. ,Giannakakis, P. E., Tselikas, N. D., & Venieris, I. (2004). Parlay-based service provision in circuit- and packet-switched telecommunications networks. *International Journal of Communication Systems. 17*(1), 63-83.

Rautela, D., Markwardt, G., & Kamel, A. (2002). Combining Open APIs (Parlay/JAIN) & Software Agents for Next Generation Mobile Services. *ISMIS02 Workshop.*

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). *SIP: Session initiation protocol*. Request for Comments (Standards Track), 3261, Internet Engineering Task Force.

Schulzrinne, H., & Rosenberg, J. (2000). The Session Initiation Protocol: Internet Centric Signaling. *IEEE Communications Magazine*, *38*(10), 134-141.

Stefano, C., Heikki, H., Oskari, K. (2000). Performance enhancing proxies for Java2 RMI over slow wireless links. *International Conference on the Practical Application of Java, pp.* 76-89.

SUN jNETx OSA Platform at www.sun.com

Tait, D., Keijzer, J. D., Goerdman, R. (2000). JAIN: A New Approach to Services in Communication Networks. *IEEE Communications Magazine*, *38*(1), 94-99.

The Parlay Group official Website, http://www.parlay.org/

Tselikas, N. D., Dellas, N. L., Koutsoloukas, E., Kapellaki, S., Prezerakos, G. N., & Venieris, I. (2007). Distributed service provision using open APIs-based middleware: "OSA/Parlay vs. JAIN" performance evaluation study. *Journal of Systems and Software* 80(5). 765-777.

Venieris, I., & Hussmann, H. (Ed.) (1998). *Intelligent Broadband Networks.* John Wiley & Sons Ltd.

Venieris, I., Zizza, F., & Magedanz, T. (Eds.) (2000). *Object Oriented Software Technologies in Telecommunications*. John Wiley, Chichester.

W3C, Web Services Description Language (WSDL) Specification, v 1.1, at http://www.w3.org/TR/wsdl

KEY TERMS

Common Object Request Broker Architecture (CORBA): A standardized framework endorsed by the OMG (Object Management Group) that specifies the interaction between software objects in a distributed computing environment.

Java APIs for Integrated Networks (JAIN): An activity within the Java Community Process, developing APIs for the creation of telephony (voice and data) services.

Network Operator: A business or organization that sells bandwidth or network access and owns a backbone network.

Next Generation Networks (NGN): A network transporting all information and services (voice, data, and all sorts of media such as video) by encapsulating these into packets.

Open Service Access / Parlay (OSA/Parlay): A technical industry consortium that specifies APIs for the network services, enabling the creation of services by organizations both inside and outside of the traditional carrier environment.

Remote Method Invocation (RMI): A Java application programming interface for performing the object equivalent of remote procedure calls.

Service Provider: A business or organization providing consumers or businesses access to Internet and telecom services.

Session Initiation Protocol (SIP): An application-layer control (signalling) protocol for creating, modifying, and terminating sessions with one or more participants.

Chapter XLVI
On the Management Performance of Networked Environments Using Web Services Technologies

Lisandro Zambenedetti Granville
Federal University of Rio Grande do Sul – Porto Alegre, Brazil

Ricardo Neisse
Federal University of Rio Grande do Sul – Porto Alegre, Brazil

Ricardo Lemos Vianna
Federal University of Rio Grande do Sul – Porto Alegre, Brazil

Tiago Fioreze
Federal University of Rio Grande do Sul – Porto Alegre, Brazil

ABSTRACT

The management of telecommunication and data networks has been based on standards defined in historical contexts quite different than the current times. As a consequence, traditional management technologies are not able to address important challenges posed by the modern infrastructures. Web Services technologies enable the proper communication of processes deployed on quite hostile environments such as the Internet. The use of Web Services for management allows the integration of low-level activities (e.g., retrieving monitoring information from gateways) with high-level business processes (e.g., creating a new product and its marketing strategy.) Despite clear advantages, Web Services-based management does not come for free; since Web Services are based on XML documents, its performance, compared with traditional management technologies, may represent an important drawback. This chapter covers the aspects of using Web Services for management focusing on the different interactions between managers and devices and the performance associated with it.

INTRODUCTION

For more than ten years the Simple Network Management Protocol (SNMP) (Case, 1990) has been used to manage networks and services. Standardized by the Internet Engineering Task Force (IETF), SNMP is not only a protocol itself but a whole management framework widely recognized and accepted by both academia and industry. Despite its evolution through the definition of SNMPv2 (Presuhn et al., 2002) and SNMPv3 (Harrington et al., 2002), the SNMP framework still has restrictions that prevent its integration with other critical disciplines, such as e-Business, e-Learning, and e-Government. For example, since SNMP traffic is normally blocked by Internet firewalls, is it not possible for different companies to exchange management information via SNMP. Also, SNMP data is encoded following rules quite different than those based on XML (eXtensible Markup Language), normally used by e-Business solutions. Thus, although SNMP could technically be used in other fields, the framework restrictions make SNMP feasible almost exclusively for network management.

Recently, the Web Services (WS) technology has emerged as an interesting and promising management alternative that could overcome some of the SNMP problems. However, since Web Services are younger than SNMP, investigations are being carried out in order to understand the difficulties and the impact in adopting Web Services for management. First investigations in this field were more focused on the network bandwidth consumption (Neisse et al., 2004) because Web Services, which are based on XML, intuitively would consume more bandwidth than SNMP, which is a binary protocol with messages supposedly smaller. Next, response time and other performance aspects such as memory consumption and processing have been investigated as well (Dreves et al., 2004) (Pavlou et al., 2004), again because in comparison to SNMP, Web Services potentially would require more memory and processing power to store and parser XML structures.

Technically, Web Services could completely replace SNMP, but that is not actually feasible because network operators would not instantaneously upgrade or replace the already deployed SNMP-enabled devices and services just because a new management framework, based on Web Services, is available. However, solely using SNMP would not allow the integration of network management-related tasks with other tasks required by those disciplines cited before. Therefore, an intermediate approach is required in order to integrate "legacy" devices and services into Web Services-based systems. That can be successfully accomplished by the use of Web Services gateways.

Gateways have been around in the network management field almost since the beginning of SNMP. CMIP (Common Management Information Protocol) to SNMP (Saydam et al., 1998) and CORBA (Common Object Request Broker Architecture) to SNMP (Aschemann et al., 1999) are examples of gateways investigated in the past whose objective was to integrate SNMP with other technologies, in this case, CMIP (OSI, 1991) defined in the ISO/OSI management framework, and CORBA (Orfali et al., 1998). The interesting point regarding gateways for Web Services integration is that they can be designed and built using different approaches, and each design approach impacts not only on the gateway building process itself, but also on the performance of the underlying managed network and associated management system.

In this chapter we present and discuss the different approaches for Web Services for management integration, namely protocol-level, object-level, and service-level gateways. We show how SNMP-enabled devices can be effectively integrated into Web Services-based systems in a feasible fashion. The discussion about the approaches for Web Services to SNMP gateways is also presented. The gateways approaches are evaluated considering a set of evaluation parameters, such as gateways' ease of use and gateways' response time.

BACKGROUND

As mentioned before, gateways in network management have been around since the SNMP beginning. Gateways based on XML and Web Services technologies, however, are more recent.

Oh *et al*. (2002) defined SNMP to XML gateways and three methods for interactive translation: DOM (Document Object Model)-based translation, HTTP (HyperText Transfer Protocol)-based translation, and SOAP (Simple Object Access Protocol)-based translation. In the DOM-based translation, an XML-based manager calls a DOM interface that resides in the gateway. Such call is then translated to SNMP operations between the gateway and the target device. With the HTTP-based translation the gateway receives XPath and XQuery expressions defined by an XML-based manager. Such expressions are then translated to SNMP requests. Finally, in the SOAP-based translation the gateway exports more sophisticated services accessed by the XML-based manager. With these services the manager can look up information with XPath or proceed with complex queries through XQuery expressions.

Strauss and Klie (2003) have proposed an SNMP to XML gateway similar to the HTTP-based translation method of Oh *et al*. The gateway accepts HTTP messages with XPath expressions in the URL. The XPath expressions are then verified to be translated to SNMP messages. DOM is then used to access the XML documents inside the gateway, reducing the data transferred between the XML-based manager and the gateway. To cope with configuration management, HTTP POST messages are translated to SNMP Set requests.

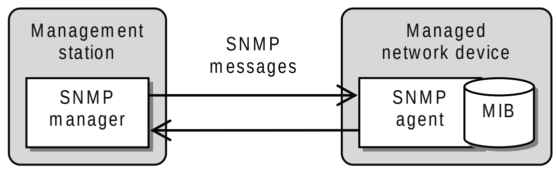
Neisse *et al*. have implemented a system that automatically creates SNMP to XML gateways given an SMI (Structure of Management Information) MIB (Management Information Base) file (Neisse et al., 2003). The gateways created retrieve information from target SNMP devices and generate XML documents sent back to an XML-based manager to be further parsed and analyzed. As the work of Strauss and Klie before, the translation is executed through the help of the `smidump` tool (Strauss 2003).

The work that has been done up to today uses just a subset of the facilities found in the WS architecture. As verified, gateways are created to access SNMP-based devices and export management information on XML documents (real WS, based on SOAP, are barely used). Besides, the WS description through WSDL (Web Services Description Language) and its registration in UDDI (Universal Description, Discovery and Integration) are not addressed in the work done so far. In addition, these investigations presented in this section do not address the diversity in building and using the developed gateways, which is critical for their proper use in real environments.

THREE APPROACHES FOR WEB SERVICES GATEWAYS

The typical management framework (which SNMP follows) is basically composed by four main elements: manager, agent, protocol, and management information base. The **manager**, from a management station, accesses the management information base (**MIB**) of a managed device contacting, via the management protocol, a protocol **agent**, usually located inside the target device. SNMP defines protocol operations to retrieve (*e.g.*, `Get`, `GetNext`) and modify (*e.g.*, `Set`) management information. Management information is defined in **MIB modules** written in plain text files according to the SMI (Structure of Management Information) (Mccloghrie 1999) specification, which is a subset of the ASN.1 (OSI 1987) language. The management information defined in a MIB module, and the way

Figure 1. An SNMP manager contacts an SNMP agent via SNMP messages to read or modify the managed network device MIB



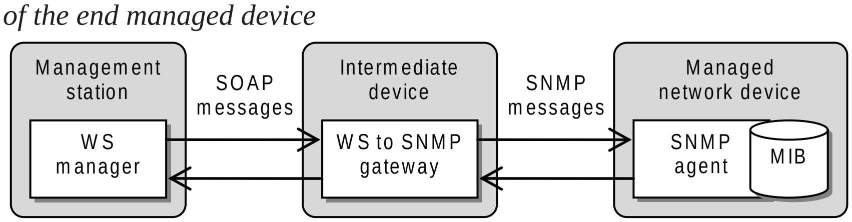
such information must be handled, defines the **management services** offered by the managed device. Figure 1 shows a conceptual view of the general SNMP framework. Further details about SNMP are available in IETF documents at http://www.ietf.org.

Two aspects of the classical management framework are important for Web Services (WS) integration. First, the SNMP framework explicitly separates the definition of the management information from the protocol that manipulates it. Second, the SNMP agent is usually just a means to access devices' MIBs through the management protocol, but it is not an active entity in the sense that it does not take any management decision based on the values of the management information. Rather, the manager is the entity responsible for analyzing the management information of a device in order to proceed with management actions, *i.e.*, management processing power is heavier at the manager side than at the agent side. The introduction of an intermediate gateway between manager and agent does not only provide the required WS to SNMP translations, but also allows balancing the processing power needed to manipulate the management information between the manager and the new introduced gateway.

In order to accomplish a WS-based management solution, the traditional SNMP manager is replaced by a WS manager, which is a WS-based client application that needs to retrieve management information from SNMP-enabled devices. A WS to SNMP gateway is placed between the new WS manager and the traditional SNMP agent in order to translate the WS manager requests to SNMP requests, and to translate back the SNMP agent replies to WS replies sent to the WS manager. In addition, WS interactions between manager and gateway are accomplished by SOAP (Simple Object Access Protocol) (Gudgin et al., 2003), which is currently the most used protocol for WS communication. Figure 2 presents the management scenario where a WS to SNMP gateway is deployed.

The creation of WS to SNMP gateways can be accomplished via different approaches, as pointed before. In the following subsections we present three main approaches. In one of them, we assume the use, at the managed device, of the Script MIB specification as an example of a more sophisticated management service. The Script MIB (Schoenwaelder et al., 2000) is a MIB module that defines management information required to allow the transferring and remote execution of

Figure 2. A WS manager contacts, via SOAP messages, a WS to SNMP gateway in order to access the SNMP agent of the end managed device



management scripts on target devices. With such MIB module a manager can send a management script to a remote device and request it to execute the transferred script. The results of the execution are then later retrieved by the manager also using the Script MIB.

Protocol-Level Gateways

Protocol-Level gateways (Schoenwaelder et al. 2003) provide communications between WS managers and SNMP agents via protocol translations, *i.e.*, SNMP messages are directly mapped to WS operations. For example, the SNMP messages `Get`, `GetNext`, and `Set` are mapped to exactly other `Get`, `GetNext`, and `Set` WS operations. In such a mapping, most data required to build up an SNMP request is provided in input arguments of the WS gateway operations. On the way back, most data carried by an SNMP response become either the WS operation results or output parameters.

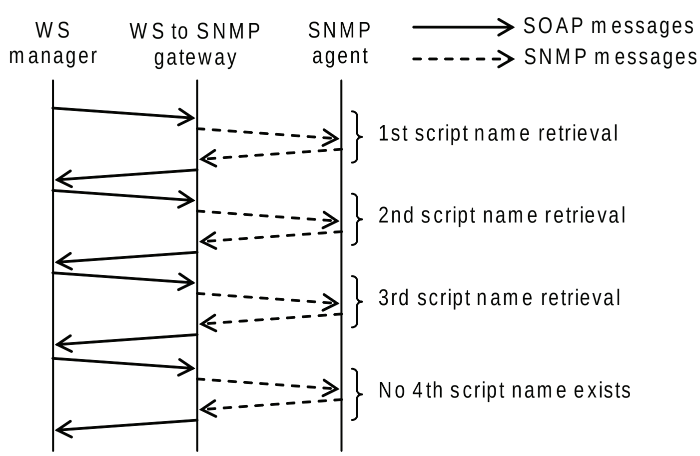
A protocol-level gateway receives SOAP requests from a WS manager and translates it to SNMP requests forwarded to the target devices. After processing the request, the target device sends back to the gateway SNMP replies that are translated to SOAP replies forwarded to the WS manager. Figure 3 shows the interactions in a scenario where a protocol-level gateway is deployed and a WS manager needs to know the name of the management scripts available in the target device. In this example, the end device hosts three management scripts. Since SNMP discovers the end of a list receiving an inadequate reply after the last element, to retrieve the name of all management scripts (*i.e.*, three names), four interactions are required.

In SNMP, each management object has an object identifier (OID), which is a sequence of integers used to address the information. For example, the `smScriptName` object of the Script MIB has OID “1.3.6.1.2.1.64.1.3.1.1.2.” OIDs are used by SNMP managers to inform SNMP agents about the internal data to be retrieved or modified. In the case of the protocol-level gateway, the WS manager still deals with OIDs to manage the target devices. Hence, OIDs are carried by SOAP messages from the WS manager and gateway, and by SNMP messages between gateway and SNMP agent.

Object-Level Gateways

Instead of translating protocol operations, object-level gateways (Neisse et al., 2004) map management information to a WS operation. For example, instead of exposing a `GetNext` operation, an object-level gateway for the Script MIB would

Figure 3. In protocol-level gateways, each SNMP message corresponds to another SOAP message



expose a `GetSmScriptName` operation to list the scripts available for execution. The `GetSmScriptName` operation is a mapping of the `smScriptName` object defined in the Script MIB.

It is important to notice that in this case the WS manager does not need to deal with SNMP OIDs anymore. The object-level gateway implementation stores the required OIDs in order to use them once the WS manager requests a device's information. Another important point is related to the retrieval of management information. In the case of the object-level gateway the gateway itself, and not the WS manager, controls the interactions with the SNMP agent required to retrieve all data from a list, builds up the SOAP reply, and sends it back to the WS manager. This interaction control, which would be performed by the manager in the protocol-level gateway, is then moved to the object-level gateway, introducing a certain level of control on the gateway side.

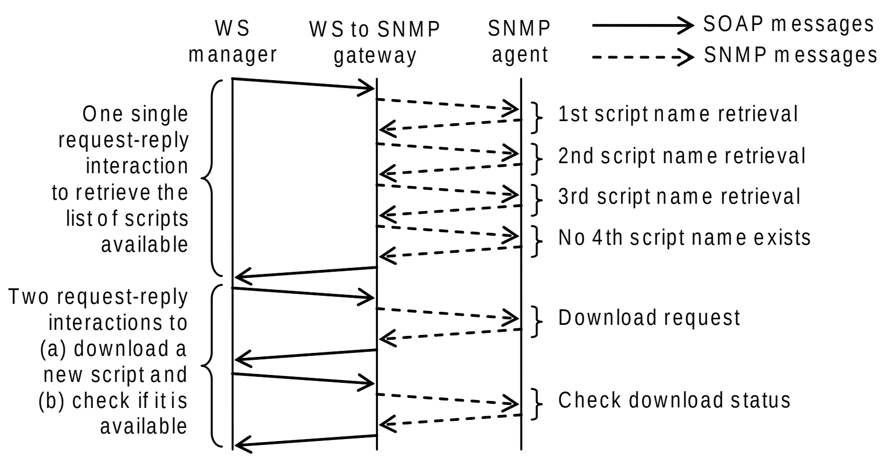
In comparison to Figure 3, Figure 4 shows that WS manager and gateway have fewer message exchanges when the object-level gateway is used. Figure 4 additionally shows the interactions needed to order the target device to download (from an external server not depicted in the figure) a management script, and associated retrieval of the download status. The example of Figure 4 deals with three different objects defined in the Script MIB: a list of scripts (`smScriptName`), an object to inform the management script location (`smScriptSource`) from where the target device must download it, and an object that reports the download status (`smScriptAdminStatus`).

Service-Level Gateways

As presented before, the set of management information defined in a MIB module ends up providing a management service available at target devices. Service-level gateways (Fioreze et al., 2005) are those that map the management services of a MIB module.

Let's consider again the Script MIB module, which defines a set of objects that need to be manipulated in order to transfer a management script from an external server to the managed device. Such objects need to be handled in a coordinated way; otherwise the download operation may fail. Although a MIB module defines the managed information in a formal way in SMI, the management service exposed by a MIB module is defined informally in the comments clauses of each object. It means that the way and order in which each Script MIB object must be manipulated to properly download a script is informally defined in the Script MIB objects comments.

Figure 4. In object-level gateways, one WS operation may be associated to several SNMP messages



Service-level gateways are built in an empirical fashion because there is no concrete element (*e.g.*, protocol or management information) able to formally define the services exposed by a MIB module. For example, the set of Script MIB objects that supports the download request, checking, and execution of a management script could be mapped to a single WS gateway operation, as presented in Figure 5.

Although building up service-level gateways is not done on top of formal basis, its usage from the WS manager is easier because such manager does not need to know neither the managed device OIDs (as in the protocol-level gateways) nor how a set of objects needs to be manipulated in order to have a management services properly working (as in the case of the object-level gateways).

PERFORMANCE OF WEB SERVICES-BASED MANAGEMENT APPROACHES

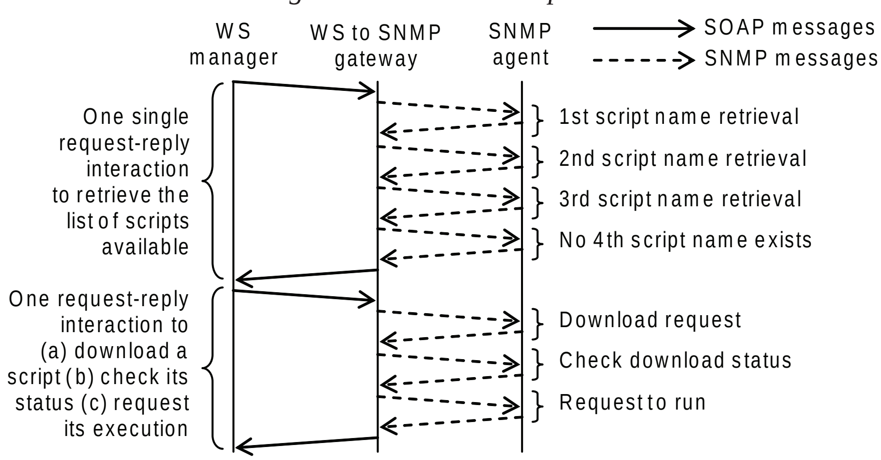
In this section we present the performance of the three WS to SNMP gateways presented before in order to characterize them in terms of network bandwidth consumption, perceived execution time, effort to maintain, and ease of use. These parameters not only allow determining the gateways' behavior and characteristics but they also indicate the management scenarios that each gateway is more suitable for.

Consumed Network Bandwidth

Network consumption is critical for management because usually the bandwidth available for network users' traffic is shared with management traffic. If management traffic grows too high, less bandwidth is available to users, which is obviously inadequate. SNMP is a quite light protocol considering the bandwidth consumed to carry the management information. SOAP, however, has messages longer than SNMP because SOAP is based on XML. This difference could lead to the conclusion that WS manager and gateway interactions consume more bandwidth than gateway and SNMP agent interactions. That is true for the protocol-level gateways, where each SNMP message corresponds to exactly another SOAP message, but it is not always true for object and service-level gateways.

Let's observe the bandwidth consumed by the three approaches considering two management scenarios. In the first one, a WS manager interacts, in different moments, with a protocol-level and

Figure 5. In service-level gateways, one WS operation may be associated to several SNMP interactions and objects in order to have a management service accomplished



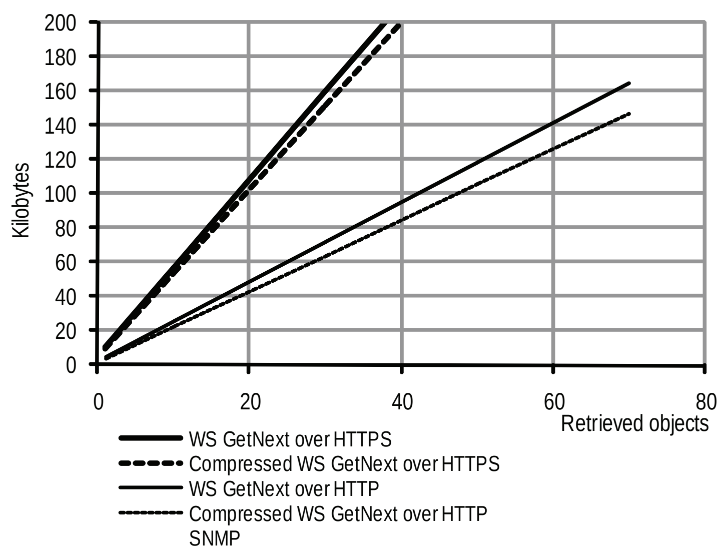
an object-level gateway in order to retrieve a list of integer objects defined only for the purpose of this test, *i.e.*, the meaning of the returned integers is not important, but the bandwidth consumed. We start with a single one integer object and progressively increase the testing list until 70 single integer objects are retrieved, and measure the network usage between the WS manager and the gateway, and between the gateway and the SNMP agent. The testing network in this first scenario is identical to the one presented in Figure 2, *i.e.*, protocol-level and object-level gateways are dedicated machines placed between manager and agent.

The graph from Figure 6 presents the network usage for the SNMP traffic and SOAP traffic running over HTTP and HTTPS (for the case of required security support). Since intuitively SOAP would consume more bandwidth than SNMP, compressed SOAP traffic has been also observed to check how much the bandwidth consumption could be decreased. As shown in the graph, SNMP traffic will always consume far less bandwidth than SOAP, even if SOAP compression is present. This situation gets even worse when the number of objects to be retrieved increase: since each SOAP message consumes more bandwidth than the associated SNMP message, the higher the number of messages, the greater the difference between SOAP and SNMP.

By replacing the protocol-level gateway for an object-level gateway, we have the bandwidth consumption presented in Figure 7. In this case, although initially SOAP traffic (over HTTP or HTTPS, and compressed and uncompressed) consumes more bandwidth than SNMP, if more objects are retrieved from the managed device SOAP traffic will eventually consumes less bandwidth than SNMP. That happens because the object-level gateway groups all SNMP retrieved information from the target device in a single SOAP reply. Thus, there is only one SOAP request and associated reply, while several SNMP interactions are present, as has been depicted in Figure 4.

By reviewing the graph from Figure 7 it is possible to conclude, for this first scenario, that uncompressed SOAP over HTTPS traffic consumes less bandwidth than SNMP if a list with more than 38 integers needs to be retrieved. Compressed SOAP over HTTPS traffic is "better" than SNMP if more than 26 integers are retrieved.

Figure 6. SNMP traffic always consumes less bandwidth than protocol-level gateway traffic, even when SOAP compression is used



Uncompressed SOAP over HTTP traffic consumes less bandwidth than SNMP after 17 integers, and compressed SOAP over HTTP after 11 integers. It is interesting to notice that for more than 68 integers, compressed SOAP over HTTPS traffic consumes less bandwidth than uncompressed SOAP over HTTP. In general, the introduction of compressing makes the SOAP traffic line angle to slow down, while the introduction of security support via HTTPS makes the SOAP traffic line initiate at a higher value.

In order to check the bandwidth consumed by service-level gateways, we put them in perspective with object-level gateways in a second scenario, where we turn back to our previous Script MIB cases. In this scenario, presented in Figure 8, we placed both gateways and SNMP agent inside the same target device. Although this configuration does not affect the bandwidth evaluation, it has important impact on the response time, to be presented in the next subsection.

In this scenario, the the WS to SNMP gateway is coded using the `nuSOAP` library for `PHP`. The target device is a Linux-based host running the Apache Web server, a local management script repository (SR), and the `Jasmin` tool. `Jasmin` (Braunschweig, 2003) is an implementation of the Script MIB developed by the Technical University of Braunschweig and NEC C&C Research Laboratories. `Jasmin` implements the Script MIB published in the RFC 2592, which was later updated by the RFC 3165. `Jasmin` supports the Java and TCL runtime engines, so that network managers can delegate Java and TCL management scripts to the `Jasmin`-enabled end device.

Figure 7. Since the object-level gateway can group SNMP retrieved information in a single SOAP reply, at some points total SOAP traffic will consume less bandwidth than SNMP traffic

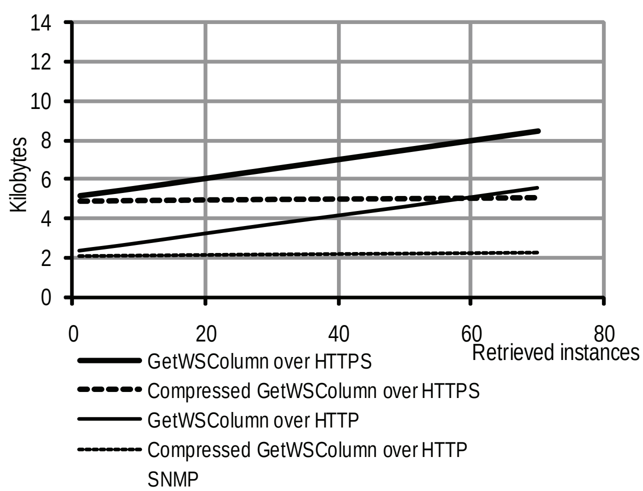
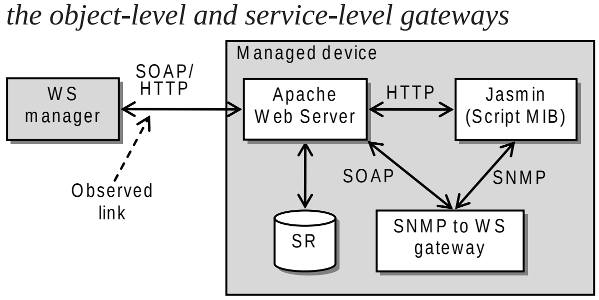


Figure 8. Scenario for the object-level and service-level gateways



The interactions on this scenarios works like this. The WS manager builds up an SOAP/HTTP request sent to the target device's Apache server, which removes the HTTP information and delivers the SOAP request to the internal WS to SNMP gateway. The gateway contacts, using SNMP, the `Jasmin` agent to request Script MIB operations. One operation is the download of a management script from the `Jasmin` agent. Since our script repository (SR) is internal to the local device, once `Jasmin` is required to download a script it locally contacts the Apache Web server in order to download, via HTTP, a management script from the SR.

In this scenario the following management actions are executed:

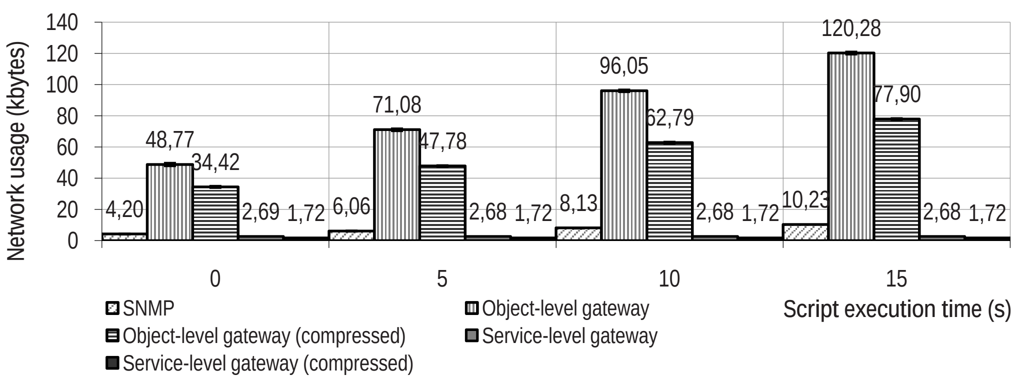
- The manager orders the SNMP agent to download a management script named `wait.jar`;
- The manager blocks until the SNMP agent finishes the script download from the local repository SR;
- The manager blocks again until `wait.jar` is ready to run;
- The manager requests the execution of `wait.jar` passing as input argument the number of seconds `wait.jar` must wait until its end (0, 5, 10, and 15 seconds are considered);
- The manager blocks again until the execution of `wait.jar` is over;
- Finally, the manager retrieves the execution results from the target device.

Although not explicitly shown on Figure 8, two different managers are in fact considered: a WS manager and an SNMP manager. The WS manager interacts with the WS to SNMP gateway, while the SNMP manager interacts directly with the `Jasmin` agent, skipping the WS to SNMP gateway. With these two managers it is possible to compare the bandwidth consumed by SOAP and SNMP to accomplish the management actions defined previously.

One important difference distinguishes the object-level and service-level gateways: in the object-level gateway the WS manager controls each step, while in the service-level gateway the gateway itself controls all steps and the WS manager just needs to order the executions of the steps to the gateway. If instead of using a WS manager one uses an SNMP manager, then the SNMP manager will behave like the WS manager in the object-level gateway, since the SNMP manager will be the one responsible for controlling the execution of the management steps.

In the second scenario, only compressed and non compressed HTTP traffic is observed; HTTPS

Figure 9. Object-level gateway, for different actions, consumes more bandwidth than SNMP, while service-level gateway, that aggregates SNMP information in a single SOAP request-reply interaction, consumes less bandwidth



is not considered because, in comparison, compression has a stronger impact than HTTPS on the final bandwidth consumed. Figure 9 shows the bandwidth consumed by SNMP in a direct SNMP manager-agent interaction, and by WS manager and WS to SNMP object-level and service-level gateways.

According to the previous testing steps, the (WS or SNMP) manager blocks in three moments: when it waits for the `wait.jar` script to be downloaded, when it waits for the script to be ready to run, and when it waits for the script to finish its execution. To implement such blocking, the SNMP manager for the Script MIB and WS manager for the object-level gateway loop until they get a proper confirmation from the target devices indicating that the script is downloaded, ready to run, or finished.

From Figure 9, SNMP consumes far less bandwidth than the object-level gateway, but more bandwidth than the service-level gateway. In the case of the service-level gateway, no loop is required in the manager because the gateway itself blocks until the script results are available. Thus, the testing loop has been moved from the WS manager to the service-level gateway, then consuming less bandwidth. Figure 9 also shows the bandwidth consumed when the SOAP/HTTP traffic is compressed. As observed, and intuitively believed, less network resources are consumed when compression is used because compressed SOAP/HTTP messages are smaller than the uncompressed ones.

It is important to note that although object-level gateway consumes less bandwidth than SNMP in the first evaluation scenario, but more bandwidth than SNMP in the second scenario, this difference is not inconsistent. In the first scenario the object-level aggregates SNMP information in a single SOAP reply because the information being retrieved was a list of elements associated to the same object (a testing list of integers), while in the second scenario the SNMP information (objects) was different from each other, preventing the object-level gateway to aggregate them. That aggregation, however, was possible in the service-level gateway, which then consumed less bandwidth.

Perceived Execution Time

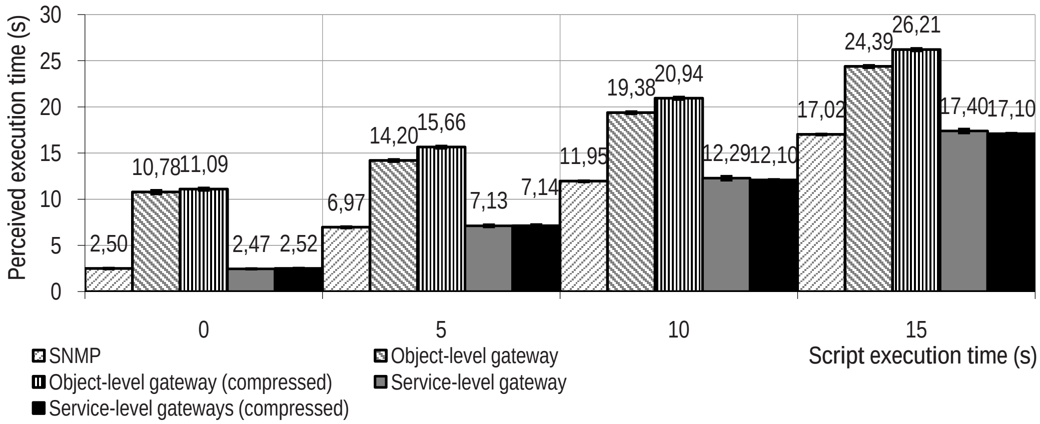
The perceived execution time is defined as the time when a manager perceives that a management task is over. That is different to the real execution time, since a management task may finish at a time, but perceived as so by the manager at a later instant. Perceived execution time is an important metric because managers take management actions normally based on the results of previous actions. If the perceived execution time is high, the management actions are delayed in the management station. Ideally, the perceived execution time should be as close as possible to the real execution time.

To check the perceived execution time, we have used the second evaluation scenario presented in the bandwidth evaluation. We have measured the execution time to download, execute, and retrieve the results of the `wait.jar` management script using, again, an SNMP manager, and two WS managers for the object-level and service-level gateways. Figure 10 presents the measured perceived execution times.

One aspect that influence in the perceived execution time is the set of software libraries used in the management system. As pointed before, we have been using the nuSOAP library to support Web Services in the PHP scripting language. In order to compare how SOAP libraries impact in the perceived execution time, we have re-implemented the service-level gateway now replacing nuSOAP by PEAR::SOAP, which is a second library for SOAP support in PHP. Figure 11 presents the results on using service-level gateways implemented using nuSOAP and PEAR::SOAP.

As can be observed, PEAR::SOAP tends to impose a greater perceived execution time, especially in the case of object-level gateways. This is

Figure 10. The execution time of actions executed through object-level gateways (with or without compression) take longer to be perceived than in the case of SNMP, while in for service-level gateway the perceived execution time is similar to SNMP



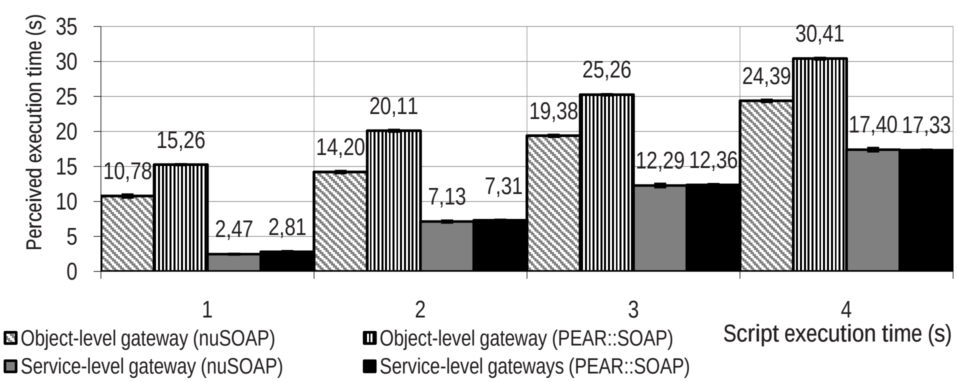
so because PEAR::SOAP has a processing delay greater than nuSOAP, then taking more time to encode and deliver the SOAP/HTTP messages to the WS manager.

Maintenance Effort and Ease of Use

WS to SNMP gateways need to be maintained in order to properly follow the changes in the target device MIB structure. Gateways need also to be updated if they are supposed to access new target devices with different MIBs than the MIBs of previous managed devices. In addition, different gateways have different ease of use from the WS manager perspective, *i.e.*, WS managers face different level of easiness in dealing with different gateways. In this subsection we check the maintenance effort and the ease of use of the three WS to SNMP gateway approaches. Concerning the ease of use aspect, we take the approach of observing the gateways from a WS manager perspective, which is different than the perspective of an SNMP manager, since SNMP managers are aware of the SNMP technology, but WS managers are not.

In the standard SNMP framework, managers require maintenance effort in order to keep in tune with the management information of target devices. SNMP manager maintenance is accomplished informing the manager about new

Figure 11. Generally, PEAR::SOAP, compared with nuSOAP, results in a greater perceived execution response time



management information using the MIB modules that describe this new information. Usually, when the manager needs to support a new device, the network human operator downloads, from the devices manufacturer Web site, MIB module files that specify the information available at the target device, and installs these modules in the manager. From this installation on, the SNMP manager is aware of the available information and can properly access and manipulate it on the target device.

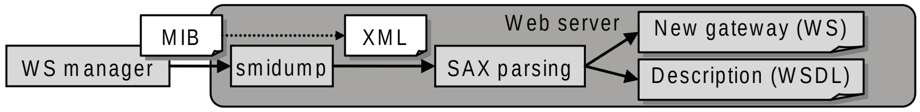
Protocol-level gateway is not aware about the information being requested to the target device: it only translates requests from SOAP to SNMP and back. In this way, if new management information is available at the target device, actually the WS manager needs to be updated, not the protocol-level gateway. Thus, once built, protocol-level gateways require no maintenance effort, considering the changes in the MIB of target devices. On the other hand, protocol-level gateways are not easy to use to WS manager. Protocol-level gateways expose operations (*e.g.*, Get, GetNext, Set) that forces the WS manager to known the SNMP information peculiarities in order to deal with them. Thus, WS managers using protocol-level gateways are forced to be SNMP aware, which is not simple. Since dealing with SNMP information (*e.g.*, OIDs) is complex, protocol-level gateways are not easy to use too.

Object-level gateways, concerning their maintenance effort, do require to be updated as soon as new information is available on target devices. This is so because the set of management information defines the WS operations exposed by the object-level gateway. If new information is included in the target device, but the gateway is not updated, the information will be unavailable for the WS manager. In this case, the WS manager will be able to access only the information that has a corresponding WS operation on the object-level gateway.

Although object-level gateways require maintenance efforts, this can be automated via supporting software able to parse and process MIB module files and generate associated gateways. Since MIB modules are defined using the SMI precise rules, one can access its content and produce outputs results convenient for specific applications, that in our case corresponds to the creation or update of object-level gateways. Figure 12 presents the architecture of an object-level gateway building system developed for this investigation. The first step in a gateway creation is the transfer of a MIB module from the WS manager to the Web server trough HTTP or HTTPS. Internally on the server, the `smidump` tool checks the received MIB module and, if no inconsistencies are found, it generates an XML temporary document. The `smidump` tool (Strauss, 2003), developed at the Technical University of Braunschweig, is a whole MIB module checking and manipulating system able to translate MIB modules to other representations, including XML. The next SAX (Simple API for XML) parsing step, in turn, takes the XML temporary document to build up the new gateway. Each object in the original MIB module is transformed in operations of the new generated WS. These operations are instrumented with code able to contact, via SNMP, a target device. The just created gateway is then stored in a standard directory in the Web server and available to be invoked just after its creation. At the same time the parsing step creates the code for the new WS, it also builds the WSDL document that describes the created WS. The WS-based manager that requested the creation of the new object-level gateway optionally informs the URL of a UDDI repository where the created gateway is registered. The original MIB module file provided by the WS manager is also stored in another standard directory for documentation purpose, as well as the intermediate XML document generated by the `smidump` tool prior to the parsing step.

With this gateway building process, new MIB modules can be easily added to a WS-based management environment. In order to let this process works properly, however, the original MIB

Figure 12. The object-level gateway building system receives a MIB module files and generates, after some intermediate steps, a new gateway and corresponding WSDL document



module files should be correctly defined. It is not rare to find MIB module files that have definition problems (mainly old MIB files, since more recent MIB modules seem to be more properly defined). For example, according to the output of the `smilint` tool, the RFC1213-MIB module has 3 errors (Fenner, 2003). In the case of finding MIB definition problems, the intended WS to SNMP gateway will not be created, and the corresponding `smidump` message describing the errors found will be sent back to the WS-based manager. It is important to notice that the gateway building system itself is not a WS, but a Web application running at the same Web server that will host the gateways just created.

In terms of ease of use, object-level gateways are more interesting than protocol-level gateways. Since SNMP management information is mapped to WS operations, SNMP details are hidden by the object-level gateway from the WS manager. Although still dealing with management information at some level, WS managers do not need to know, for example, how to manipulate SNMP OIDs in order to retrieve data from the target devices. Given that the object-level building process presented before generates an associated WSDL document, WS managers can be aware of the new management information whenever the target device MIB is updated (and an associated gateway built). The important point is that the WS manager is not forced to deal with MIB modules files, *i.e.*, WS managers do not need to be SNMP aware anymore.

Service-level gateways, compared with protocol-level and object-level gateways, require far more efforts to be maintained. As presented before, service-level gateways are not built upon formal basis (protocol or MIB module definition), but rather based on the services exposed by a MIB module which are defined in the MIB module comment clauses. It means that for every new MIB module available that requires a service-level gateway, such gateway will be developed manually checking the MIB module content to produce a set of WS operations. The developer (or the designer) of service-level gateways is the one responsible for mapping the original MIB module into the gateway operations. That means that two different developers (or designers) will probably end up with different gateways for the same original MIB module.

Since the services are not formally defined, there is no current tool able to, given a MIB module, automatically produce a new service-level gateway, as it was the case for object-level gateways presented before. Thus, for each new MIB module available, the developer has to built the associated service-level gateways, which is a slower process if we compare with the object-level gateways (that are automatically created via a building system) and protocol-level gateways (that need to be implemented just once, regardless the changes in the MIBs of target managed devices).

However, regarding ease of use, service-level gateways are quite interesting because besides hiding SNMP details from the WS manager (as object-level gateways do), service-level gateways also hide the steps required to accomplish a specific management service defined in a MIB module. While in the object-level gateway the WS manager needs to know what WS operations must be called and in which order, in the service-level gateway the WS manager only needs to invoke a WS operation that corresponds to a whole management service. Surely, this ease of use depends on the quality of the mapping of the MIB service to the service-level gateway, but in general it is safe to say that even for bad mappings the service exposed by a service-level gateway is easier to use than the corresponding whole set of operations exposed by an object-level gateway.

In order to summarize the evaluation of the WS to SNMP gateways, table 1 puts the three approaches in perspective regarding the evaluation parameters consumed bandwidth, perceived execution time, maintenance effort, and ease of use.

CONCLUSION

In this chapter we have evaluated three different approaches for building Web Services (WS) to SNMP gateways. Such gateways are required for the integration of traditional network devices into WS-based systems, such as systems for e-commerce and e-government. Protocol-level gateways directly map SNMP messages to SOAP messages, while object-level gateways hide SNMP details from WS managers offering WS operations mapped from MIB modules. Finally, service-level gateways go further and expose management services instead of management information (like object-level gateways) or protocol mapping (like protocol-level gateways). However, service-level gateways are more difficult to maintain because they require the human interpretation of MIB modules to build up a new version of the gateways when the set of management information is changed. Object-level gateways, on the other hand, can be automatically built given a MIB module as input argument, while protocol-level gateways do not require maintenance efforts at all because their operation is not affected by changes in the management information available at the target devices.

Protocol-level gateways are interesting in management environment where the set of management information changes quite frequently. Although they consume far more bandwidth than SNMP, protocol-level gateways may enable the communication of manager and agents located in different administrative domains because SOAP traffic tends to cross Internet firewalls easier than SNMP. However, the WS manager must be aware of the SNMP peculiarities in order to interact with the protocol-level gateway, which is an important restriction when dealing with systems of broader disciplines such as e-business.

Object-level gateways are ideal when the amount of information to be retrieved from managed devices is high. Since these gateways are able to group SNMP information in a single SOAP

Table 1. Summary on WS to SNMP gateways evolution

Gateway	Bandwidth	Execution time	Maintenance	Ease of use
Protocol-level	Gateway with highest bandwidth consumption	Takes more time than SNMP to learn that a task is over	After being built, requires no maintenance at all	Difficult to use because forces WS managers to be SNMP aware
Object-level	Consumes less bandwidth than SNMP for a large number of objects	Takes far more time than SNMP to learn that a task is over	Easy to maintain given a gateway building system based on MIB modules	Hides SNMP details, but WS managers need to know how to use the service
Service-level	Consumes less bandwidth than SNMP and other gateways	Learns faster than SNMP that a task is over	Difficult to maintain and requires human interpretation of MIB services	Easy to use because service details are hidden from the WS manager

reply, object-level gateways tend to consume less bandwidth than SNMP for a large number of information. That enables protocol-level gateways to be used, for instance, in monitoring system composed of measurement elements deployed in networks located in different administrative domains, and bunches of collected data needs to be transferred from time to time to a central Web Services based analysis tool. Since the maintenance and ease of use of object-level gateways are not complex, the availability of new different management information on the managed network does not lead to complex updates of gateways and WS managers.

Finally, service-level gateways, besides consuming less bandwidth than SNMP and presenting response time close enough to SNMP, has the advantage of being easier to use. Service-level gateways are ideal for integrating SNMP into non network management systems, because these gateways not only hides the SNMP complexities but also exposes management services in a simpler way than the previous gateways. If we do not consider the maintenance effort, the results presented in this chapter help us to argue in favor of a service-oriented SNMP/WS integration, instead of information or protocol-oriented integration. The presented service-level gateway is easier to use (from the manager perspective), consumes less bandwidth than the other gateways, and presents a response time quite close to SNMP. It is important to highlight that these conclusions are based on the case of the Script MIB. However, although the SNMP framework is well known and accepted, designing WS management solutions taking the SNMP framework as a starting point is not a proper approach in modern times. Rather, services must be considered in the first place and SNMP "details" should be treated only when the protocol and related information issues begin to be relevant.

A WS feature that is often cited is that more complex and richer WS can be built from the combination and reuse of other simpler WS. This feature could be used to integrate network management services, or, more interestingly, to create new services that could not be easily built today except through heavy coding efforts. Integrating simpler WS to form richer network management WS is one of the interesting challenges to be addressed. Web Services coordination, orchestration, and choreography applied to the network management field are certainly aspects that must be investigated and considered in the network integration and management field. This is a feasible way to achieve a real and long awaited integration between network management and other Web Services-based disciplines.

REFERENCES

Aschemann, G., Mohr, T., & Ruppert, M. (1999). Integration of SNMP into a CORBA and Web-based Management Environment. In Steinmetz (Ed.), *KiVS'99* (pp. 210-221).

Braunschweig, T., & Europe, N. C. (2003). *Jasmin: A Script-MIB Implementation*. From http://www.ibr.cs.tu-bs.de/projects/jasmin

Case, J., Fedor, M., Schoffstall, M., & Davin, J. (1990). A Simple Network Management Protocol (SNMP). *Internet Engineering Task Force, Request for Comments: 1157 (STD: 15)*.

Drevers, T., va de Meent, R., & Pras, A. (2004). On the standardization of Web Services management operations. In *10th Open European Summer School and IFIP WG6.3 Workshop* (pp. 143-150).

Fenner, B. (2003). *Smilint output for all MIBs published in RFCs and Internet Drafts.* From http://www.icir.org/fenner/mibs/htmllint

Fioreze, T., Granville, L. Z., Almeida, M. J., & Tarouco, L. R. (2005). Comparing Web Services with SNMP in a Management by Delegation Environment. In *IFIP/IEEE International Sym-*

posium on Integrated Network Management (pp. 601-614).

Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.-J., & Nielsen, H. F. (2003). SOAP Version 1.2 Part 1: Messaging Framework. *W3C Recommendation*, from http://www.w3.org/TR/2003/REC-soap12-part1-20030624/

Harrington, D., Presuhn, R., & Wijnen, B. (2002). An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. *Internet Engineering Task Force, Rquest for Comments: 3411 (STD: 62).*

ISO/IEC 9596 (1991). *Common Management Information Protocol Specification*.

Mccloghrie, K., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., & Waldbusser, S. (1999). Structure of Management Information Version 2 (SMIv2). *Internet Engineering Task Force, Request for Comments: 2578 (STD: 58).*

Neisse, R., Granville, L. Z., Almeida, M. J., & Tarouco, L. (2003). A Dynamic SNMP to XML Proxy Solution. In *8th IFIP/IEEE International Symposium on Integrated Network Management* (pp 481-484).

Neisse, R., Vianna, R. L., Granville, L. Z., Almeida, M. J. B., & Tarouco, L. M. R. (2004). Implementation and Bandwidth Consumption Evaluation of SNMP to Web Sevices Gateways. In *IEEE/IFIP Network Operations and Management Symposium* (pp. 715-728).

Oh, Y. J., Ju, H. T., Choi, M. J., & Hong, J. W. (2002). Interaction Translation Methods for XML/SNMP Gateway. In *13h IFIP/IEEE International Workshop on Distributed Systems: Operations and Management* (pp. 54-65).

Orfali, R., & Harkey, D. (1998). *Client/Server Programming with Java and CORBA*. John Wiley and Sons, New York, USA, 2nd edition.

OSI (1987). Specification of Abstract Syntax Notation One (ASN.1). *International Standard* 8824.

Pavlou, G., Flegkas, P., Gouveris, S., & Liotta, A. (2004). On Management Technologies and the Potential of Web Services. *IEEE Communications*, *42*(7).

Presuhn, R., Case, J., Mccloghrie, K., Rose, M., & Waldbusser, S. (2002). Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP). *Internet Engineering Task Force. Request for Comments: 3416 (STD: 62).*

Saydam, T., & Sirsikar, R. (1998). Design of CMIP-SNMPv2 Proxy Gateway for Network Management Interoperability. *Journal of Networks and Systems Management*, *6*(2).

Schoenwaelder, J., Quittek, J., & Kappler, C. (2000). Building Distributed Management Applications with the IETF Script MIB. *IEEE Journal on Selected Areas in Communications*, *18*(5), 702-714.

Schönwälder, J., Pras, A., & Martin-Flatin, J. P. (2003). On the Future of Internet Management Technologies. *IEEE Communications Magazine*, *41*(10), 90-97.

Strauss, F. (2003). *libsmi - A Library to Access SMI MIB Information*. From http://www.ibr.cs.tu-bs.de/projects/libsmi/

Strauss, F., & Klie, T. (2003). Towards XML Oriented Internet Management. In *8th IFIP/IEEE International Symposium on Integrated Network Management* (pp. 505-518).

KEY TERMS

Integrated Network Management: The study of integrating network management operations among each other or with other external disciplines.

Network Management: A computer science discipline that investigates activities, methods, procedures, and tools to operate, administrate, maintain, and provide networking infrastructures.

Protocol Evaluation: The observation of a network protocol in terms of its behavior and the impact associated to it over the communication infrastructure.

Protocol Gateway: A software entity able to translate protocol messages, operations, and/or services from one specific protocol to another specific protocol.

SNMP: Stands for *Simple Network Management Protocol* and is the *de facto* standard network management protocol for TCP/IP networks.

SOAP: Stands for *Simple Object Access Protocol* and is the basis protocol to support Web Services communications over the Internet.

Web Services: A set of technologies, mostly based on XML, that allows inter-process communication among machines located on the Web.

Chapter XLVII
Automated Fault Management in Wireless Networks

Raquel Barco
University of Málaga, Spain

Pedro Lázaro
University of Málaga, Spain

ABSTRACT

In the near future, several radio access technologies will coexist in Beyond 3G mobile networks (B3G) and they will be eventually transformed into one seamless global communication infrastructure. Self-managing systems (i.e. those that self-configure, self-protect, self-heal and self-optimize) are the solution to tackle the high complexity inherent to these networks. In this context, this chapter proposes a system for automated fault management in the Radio Access Network (RAN) of wireless systems. The chapter presents some basic definitions and describes how fault management is performed in current mobile communication networks. Some methods proposed for auto-diagnosis, which is the most complex task in fault management, are also discussed in this chapter. The presented systems incorporate Key Performance Indicators (KPIs) to identify the cause of the network malfunction.

INTRODUCTION

There is no doubt that during the last decade mobile communications have played an increasingly important role in the telecommunication business, and it will continue to do so in the years to come. In the last years, 3G networks, called Universal Mobile Telecommunications Service (UMTS) networks in Europe, have started to be deployed throughout the world. In the near future, thanks to 3G, mobile internet-services are expected to be available "anywhere and anytime''. Users will surf the Web, check the email, download files or have real time videoconference, in a shopping mall, the airport, the city center or their homes. Beyond 3G mobile networks (B3G) (Jamalipour, 2005) will comprise a set of interrelated and rapidly growing wireless networks, applications which will require increasing bandwidth, and users who will demand high quality of service at low cost, all within a limited spectrum allocation. In these networks, the highly complex and heterogeneous Radio Access Network[1] (RAN) will be composed of different technologies, such as GSM, UMTS and WLAN.

Until now, most operational tasks have been manually performed, requiring dedicated staff, with subsequent, inflexibility and delay of response. However, network operators are currently showing a growing level of interest in automating most network management activities. This has stimulated intense research activities in the field of self-managing networks (Pras, 2007; Kephart, 2003; Strassner, 2004). In this context, the self-managing property refers to the capability of the network to self-configure, self-protect, self-heal and self-optimize. All these issues have been the main driver behind recent studies dealing with automation and optimization of cellular networks (Halonen, 2003; Johnson, 2004; Lempiäinen, 2001; Laiho, 2002a).

In a mature cellular network that has undergone most of its site roll-out, the major cost is associated to the operation of the network. As the network consists of a high number of pieces of equipment that are distributed across the entire country, maintaining and operating this large and technically complicated system is a difficult task that requires operator personnel around the clock in several regional offices. For example, a GSM network in a typical European country may consist of about 10.000 sites. Due to the large size of the networks, it is common that some of the deployed pieces of equipment do not work as planned. The consequence of such problem is poor end-user service. As in most countries several operators are competing for subscribers, it is imperative to rectify such occurrences because otherwise users will be dissatisfied with the service and thus will likely switch to competing network operators. Hence, fault management, also called troubleshooting (TS), is a key aspect of operating a cellular system in a competitive environment. As the RAN of cellular systems is by far the biggest part of the network, most of the TS activities are focused on this area.

TS comprises the isolation of faulty cells (*fault detection*), the identification of malfunctions (*diagnosis*) and the proposal and deployment of healing actions (*solution deployment* / *fault recovery*). Currently, in most cellular networks, TS is a manual process, accomplished by RAN experts. These engineers are applying a series of customized checking routines on a daily basis in order to identify the cause of the problem. During the procedure, several applications and databases have to be queried to analyze performance indicators, cell configurations and alarms. The speed in identifying faults depends on the level of expertise of the troubleshooter, the type of information available and the quality of the tools displaying relevant pieces of information. This means that, in addition to a good understanding of the possible causes of the problems, a very good understanding of the tools available to access the sources of information is also required.

In this scenario, the benefits of automating TS are numerous. With the help of an automated TS tool, the time required to identify the reason for a fault causing a problem is greatly reduced. This means that network performance is enhanced as the downtime and the time with reduced quality of service (QoS) are significantly limited. In addition, by automating the TS process, fewer personnel and, thus, fewer operational costs are necessary to maintain a network of a given size. The TS process is de-skilled as the majority of problems can be rectified with the help of the automated TS tool. Then, the knowledge of highly experienced staff, which is released from the TS work, can be utilized for other aspects of network optimization, thereby further increasing network performance. One additional benefit is that the knowledge in TS can be stored in the TS tool, therefore not being dependent on the staff working for the company at any time. In conclusion, the gains achieved thanks to automated TS for an operator are significant as fewer personnel with a lower skill level can solve more network problems in less time.

The main aim of this chapter is to describe an automated fault management system for the RAN segment of wireless networks. In particular, this chapter will be focused on auto-diagnosis, which is the most complex task in automated fault management. Diagnosis in the RAN, up to now, has received little attention in the existing literature, despite the huge interest of network operators and manufacturers of equipment (Altman, 2006). However, automatic diagnosis has been extensively studied in other fields, such as diagnosis of diseases in medicine (Ng, 2000), troubleshooting of printer failures (Heckerman, 1995) and diagnosis in the core of communication networks (Steinder, 2004). Nevertheless, diagnosis in the RAN of cellular networks has some distinctive characteristics, such as the continuous nature of most symptoms, the existence of logical faults (e.g. a wrong configuration, not related to a physical piece of equipment), the constant changes in the network (new pieces of equipment and network configuration) and in the environment (e.g. new buildings, obstacle disappearance, ...), the random nature of radio propagation, the lack of test and training examples, etc. All these peculiarities make automatic diagnosis techniques used in other application domains not directly applicable to the RAN of wireless networks.

This chapter is organized as follows. Section 2 explains some concepts and presents a survey on the state of the art in automated fault management in the RAN. Section 3 describes how fault management is currently performed in existing cellular networks and it proposes some solutions for the design of an automated troubleshooting system in the RAN of any wireless network. In addition, it presents some study cases carried out in live wireless networks. Section 4 discusses some implementation issues and presents some future trends. Finally, Section 5 summarizes the main conclusions.

BACKGROUND

Definitions

The first step in troubleshooting is the detection, which is the identification of the cells with problems. A *problem* is a situation occurring in a cell that has a degrading impact on the service. Every operator uses a different method to identify the problematic cells, which can be based on different performance indicators, e.g. dropped calls, access failures, congestion, etc. The most severe problem for mobile network operators are cells experiencing a high number of dropped calls because a dropped call has a very negative impact on the service offered to the end user. In that sense, the dropped call rate (DCR) is a good indicator of the quality of the cell.

Once the cells with problems are isolated, a diagnosis of the cause of the problems should be done for each problematic cell. A *cause* or *fault* is the defective behavior of some logical or physical component in the cell that provokes failures and generates a high DCR, e.g. bad parameter value, hardware fault, etc. A *symptom* is a performance indicator or an alarm whose observed value might be used to identify a fault, e.g. the number of handovers due to interference. The most important performance indicators are named Key Performance Indicators (KPIs). The aim of the diagnosis system is to identify the cause of a problem based on the values of some symptoms.

State of the Art

Preliminary studies related to automated fault management in the RAN of wireless networks have been focused on methods to achieve efficient visualization of the network performance. Thanks to those visualization tools, fault detection is performed more easily.

Several studies have been carried out in the field of Fault Detection (FD) (Laiho, 2005; Barreto, 2005; Lehtimäki, 2005b; Laiho, 2002; Hoglund, 2000; Lehtimäki, 2005a; Barreto, 2004), which is the first step in fault management. Most FD methods consist in building models for the normal behavior of the system. The deviations of the available measurement variables from the normal behavior can then be detected with some type of abnormality detector.

Amongst troubleshooting tasks, diagnosis of faults is the most complex and time-consuming one. However, very few references can be found on automatic diagnosis in the RAN of wireless networks. Some techniques to automate diagnosis have been proposed in other application domains such as the diagnosis of diseases in medicine (Andreassen, 1987; Heckerman, 1992; Pang, 2004), the troubleshooting of printer failures (Heckerman, 1995; Skaanning, 2000b; Skaanning 2000a), the diagnosis of faults in satellite communication systems (Lázaro, 2002; Cayrac, 1994) and fault identification in the core network (CN) part of communication networks (Katzela, 1995; Steinder, 2004; Deng, 1993). The latter is the closest field to the RAN of cellular networks, although it presents significant differences. On the one hand, in CNs it is very important to represent the dependencies among communication system entities because, due to the highly interconnected architecture, a failure in an entity may have a large impact on other system entities. On the other hand, in general, in CNs the only type of symptoms are the alarms related to the network entities. In the RAN part of cellular systems, this interconnection is not as strong and alarms are not the only symptoms of malfunction, due to two main reasons. Firstly, most faults are not related to a physical component, but to a poor network planning or an incorrect parameter setting. Thus, modelling the interactions among entities is not a requirement like in other communication networks because, in most cases, faults are not related to pieces of equipment which are physically connected. Secondly, although alarms play a very important role in identifying faults in specific pieces of equipment, they do not provide conclusive information in order to isolate configuration problems. Furthermore, the fact that performance indicators in the RAN domain are continuous, instead of binary like alarms, generates a new difficulty in the modelling, which is nonexistent in the CN domain.

Research studies on automation of diagnosis in the RAN of cellular networks have been focused on alarm correlation (Wietgrefe, 1997; Wietgrefe, 2002). In current cellular networks, most systems are semi-intelligent and generate alarm messages when errors occur. The abstraction level of these alarm messages is normally very low, leading to a high number of alarms for any single cause. For example, when a link fails, up to 100 or more alarms are generated and passed to the Network Management System (NMS). Those alarms should be converted into a minimum number of alarms which clearly pinpoint the breakdown of a link. Alarm correlation (Jakobson, 1993; Yemini, 1996) consists in the conceptual interpretation of multiple alarms, so that new meanings are assigned to the original alarms. It is a process that involves different tasks: reduction of multiple occurrences of an alarm into a single alarm, inhibition of low priority alarms in the presence of higher priority alarms, substitution of a specific set of correlated alarms by a new one, etc. This way, alarm correlation systems are required to filter and condense the incoming alarms to meaningful high level alarms in order to avoid overloading the operators. Although alarm correlation can be considered a first step in the diagnosis of faults, alarms do not provide enough information to identify the cause of problems, specially if the possible causes are not only faults in pieces of equipment. Even after alarm correlation, the number of triggered alarms for a single cause is normally very high. In addition, the same alarms may be triggered by different causes. Due to these reasons, in order to achieve a conclusive diagnosis, it is important that not only alarms are taken into account, but also network performance indicators.

In (Barco, 2002b; Barco, 2002a) an auto-diagnosis system for the RAN of cellular networks was proposed. Discrete Bayesian Networks (BN) (Pearl, 1988; Jensen, 2001a) was the technique used to diagnose the cause of the problems in a faulty cell. The system analysed the value of the main KPIs and alarms of a faulty cell in order to carry out the diagnosis. The drawback of such a system is that the number of parameters to be specified is so large that it is impossible for a TS expert to be accurate in his definition. Taking into account this inherent inaccuracy in parameter elicitation, a method was proposed in (Barco, 2009a) in order to enhance the diagnosis accuracy[2] of the auto-diagnosis system. In (Barco, 2005), an alternative system was proposed, which modelled KPIs as continuous variables instead of as discrete ones. Results obtained with the discrete and the continuous models were compared in (Barco, 2008b).

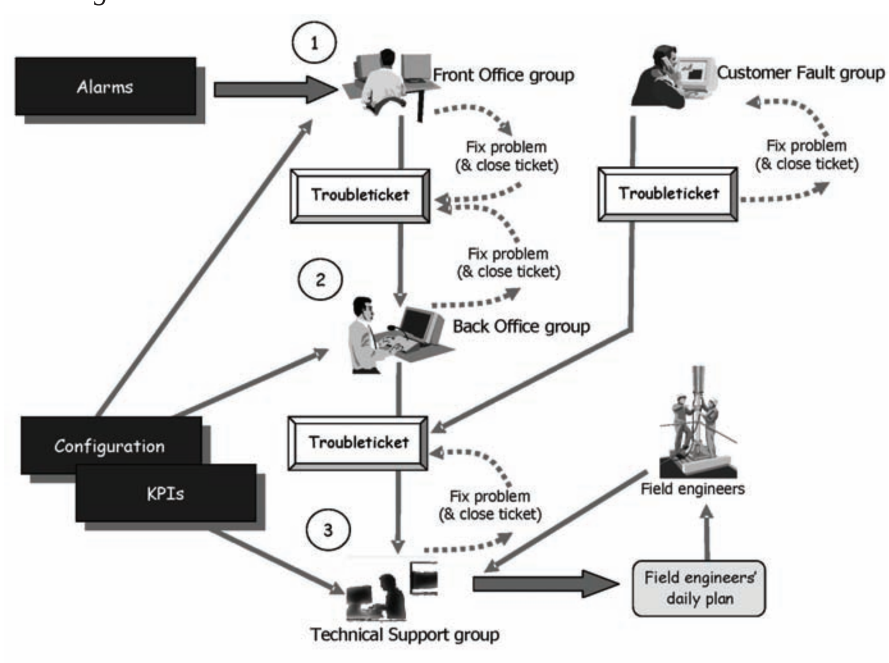
AUTOMATED FAULT MANAGEMENT

Fault Management in Current Wireless Networks

Currently, troubleshooting is mainly a manual process performed by different teams. Each network operator follows its own approach, although Figure 1 is in line with TS in most networks.

TS is normally related to a *Trouble Ticket* (TT) system. Trouble ticket systems are commonly used in telecommunication networks to assist in fault management (Melchiors, 2000; Madruga, 1994). When a fault is investigated, a TT reflects the problem status, which is the fault description, the steps followed so far to solve it out and the identification of the faulty equipment in case the problem is deemed as a hardware fault. Thus, TS begins with the documentation of a trouble in a TT. Subsequently, the trouble ticket system receives as inputs the reports generated by the different teams in Figure 1. Management of these reports implies the assignment of the tickets at each moment to

Figure 1. Fault management in current mobile communication networks



the most adequate team to deal with them. The trouble ticket may pass through several hands and undergo different levels of processing with respect to priority until it is closed, i.e. the problem is solved. A TT system is deployed as a large database, which can be queried by the user, using criteria like time constraints or identifiers (ID), such as cell ID and site ID. After a query, all the cases related to the specified cell within a given time period are shown. The entries are normally in "free text'', almost like a "virtual log-book'' that everyone uses to annotate the actions taken and observations made for the cell. Hence, in difficult cases many people from various departments have looked into the cell and potentially applied some changes. For example a case might involve changing parameters, swapping hardware, re-tuning the cell, taking more measurements, re-tuning neighbouring cells, etc. Thus, several notes are normally added to such a case.

As shown in Figure 1, TS procedure may involve several layers, depending on the complexity of the case. The Front Office group (FOG) is the first such layer. This team is responsible for dealing with alarms generated by the network. Typically, the number of triggered alarms may be up to 100000 per week for the whole network. Alarm reduction techniques reduce this large number of alarms so that the FOG is presented with the important alarms, leaving about 600 alarms per week. Then, the FOG raises TTs, i.e. reports related to those alarms. In addition, the FOG carries out a preliminary analysis to identify the cause of the problem. Sometimes this group solves the problem. In that case, they update the TT with a description of the performed actions and "close the TT''. In more complex cases, they send the TT to the second layer for further investigation, including in the TT a description of the steps that were followed.

The second TS layer is the Back Office group (BOG), which looks into faults based mainly on short-term statistics. The BOG staff initiate a deeper analysis to identify the cause of the problem and they update the TT with the executed actions. If, as a consequence, the problem is solved, they close the TT. If they are not able to solve the problem, they reassign the TT to a more specialized group, the Technical Support group (TSG). This new group may be composed of troubleshooting experts of the operator itself or external experts, e.g. from a equipment manufacturer.

Therefore, the TSG is the following layer in the procedure. This team often uses scripts to generate a list of "worst performing cells'' (i.e. they use fault detection tools) and takes this as a starting point for their work, focusing on the most serious cases. Then, they look further into TTs raised by the FOG and the BOG and raise new TTs with the results of their analysis. They can solve parameter related problems, but they often need to involve field engineers for problems related to hardware on the site. Field engineers travel to the base station sites and fix hardware problems or any other problem requiring on-site personnel. Each day the field engineers receive a new plan containing the list of sites they have to visit and the cases they have to investigate on site. If the problem is solved at the site, the TSG closes the TT.

Finally, a minor part of TTs are raised by the Customer Faults group (CFG) who may receive customer complaints from call centers, management, engineering staff, etc. They can either solve the problem themselves or send the TT to a more adequate group.

Architecture of Automated Fault Management System

Figure 2 shows the *Automated Troubleshooting System* (TSS) and its interfaces. The *Fault Detection* subsystem (FDS) provides the *Auto-Diagnosis System* (ADS) with the list of faulty cells to be diagnosed. The ADS requires a *diagnosis model* on which its reasoning mechanisms are based. The subsystem named *model definition* is in charge of building the diagnosis model to be used by the

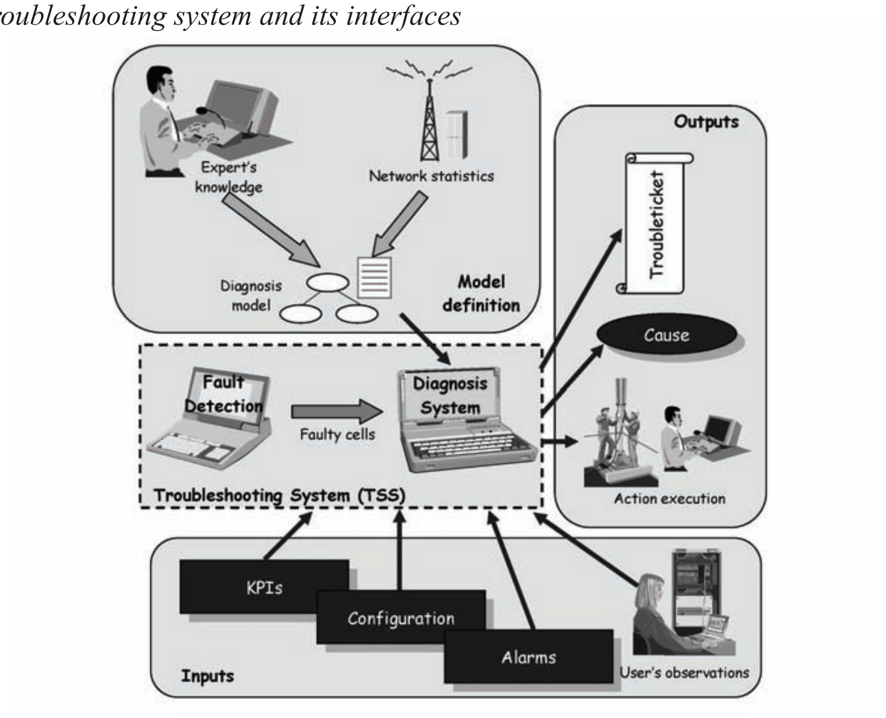
ADS. Diagnosis models can be built based either on the expertise of human troubleshooters or on statistics from the network. Those statistics are normally saved in the NMS. The inputs to the ADS are configuration parameters, alarms and KPIs for each of the faulty cells. The NMS contains historical databases containing the values of all those inputs. In addition, the ADS may also require some inputs directly requested to the user, which may be related to observations that are not in the NMS, e.g. whether the day was rainy on the day the fault occurred. The output of the ADS is a diagnosis on the fault that is causing the problems in each malfunctioning cell. In addition, the ADS proposes a list of actions, ranked by their efficiency (see Section 3.3), to be sequentially executed until the problem is solved. These actions may be just changing a configuration parameter from a remote terminal or may involve sending personnel to a site to replace a faulty piece of equipment. A TSS may even execute software related repair actions. Nevertheless, normally, operators prefer that the TSS only proposes the actions, but the final decision be handled by a human expert (the TSS in this case acts as a so called *decision support system*). Finally, the ADS is also intended to generate a report about the diagnosed cause and the steps carried out in order to recover from the fault (i.e. it will be integrated with the trouble ticket system).

Decision-Theoretic Troubleshooting

At any stage of the TS process, there are many possible pieces of evidence or tests (*observations*), and repairs (*actions*) that can be collected or applied, respectively. Because these operations are expensive in terms of time and money, it is desirable to generate a sequence of *steps* (observations or actions) that, whilst minimising costs, results in a functioning network. This is related to *decision-theoretic troubleshooting* (Heckerman, 1995; Breese, 1996; Jensen, 2001b), which considers that the aim of TS is to solve the problems, not just determining the cause.

A very simple TS procedure collects all observations. Then, it makes reasoning based on the observations, ranking the possible causes

Figure 2. Troubleshooting system and its interfaces



according to their probabilities, and it performs the repair action associated with the most probable cause. If after executing this repair action it is observed that the most probable cause is not the actual cause, the repair action associated to the following cause in the list of ranked causes is carried out. This is not an optimum procedure because all observations are collected, when probably not all of them are required to find out the cause. This is especially important in the medical domain, where some tests may be expensive and invasive for the patient. Furthermore, sometimes it could be more efficient to perform an action different to the one associated to the most probable cause. For example, suppose that the most probable cause was a hardware fault. In that case, it is probably more cost-efficient to make a simple parameter change to check whether the problem was a configuration problem than replacing the faulty hardware unit. Only if the parameter change fails to solve the problem, someone will be sent to the site to change the problematic hardware unit, which is always a very expensive and time-consuming action.

An optimum procedure[3] would arrange the steps (observations and actions) in a sequence minimising the cost (money and time). The actions would be ranked according to their efficiency, where the *efficiency* of an action is defined as the ratio between the probability of the action solving the problem and the cost of that action. At each moment, a greedy algorithm would calculate the expected cost of the sequence of steps so far executed together with each new candidate observation and action. The chosen step would then be the one with the lowest expected cost. Following this algorithm, in general, some actions will be performed before collecting all observations or before more probable, but more expensive actions. Then, very probably the problem will be solved by one of these actions, saving time and money. For example, if after collecting some observations, the probabilities of faulty hardware and bad configuration are very high, a parameter change will be proposed by the algorithm (even if the hardware fault is the most probable cause) and if the problem is solved, it will not be required to collect the remaining observations or perform the alternative actions.

In the problem under study, the RAN of wireless networks, normally all observations (KPIs and alarms) are available in the NMS and the cost of collecting them is negligible. In addition, out of the three TS tasks (fault detection, diagnosis and repair), diagnosis is the most difficult and time consuming one and no information about it can be found in present literature. Hence, the rest of this chapter is focused on diagnosis. It is supposed that there is a Fault Detection module which identifies the problematic cells prior to diagnosis. Under these assumptions, the optimal repair sequence for a faulty cell is given by the following algorithm:

1. Collect all available observations for the cell under study
2. Given that the cell is malfunctioning, compute the probabilities of the causes, $P(C_i|E)$
3. Execute the action A_i with the highest efficiency $\varepsilon(A_i|E)$:

$$\varepsilon\left(A_i|E\right)=\frac{P\left(A_i=yes|E\right)}{C_{A_i}\left(E\right)} \quad (1)$$

   where $P(A_i = yes|E)$ is the probability of the action A_i solving the problem given the evidence compiled so far and the result of previous actions, E, and $C_{A_i}\left(E\right)$ is the cost of performing the action A_i, which may depend on E. It is assumed that related to each cause C_i there is a unique action A_i that solves the problem. Hence, $P(A_i = yes|E)$ is also the probability of the cause C_i, $P(C_i|E)$.
4. If the action solved the problem, then terminate. Otherwise, go to step 2.

In this algorithm, the required pieces of information are the posterior probabilities of the causes given the evidence, $P(C_i|E)$, and the costs of the related actions, $C_{A_i}(E)$. These costs can be easily provided by operators of cellular networks. Therefore, cost will not further explicitly appear in this chapter, instead it will focus on calculating the probabilities of the causes, $P(C_i|E)$.

Bayesian Modelling of Fault Diagnosis

Two components of the auto-diagnosis system may be distinguished: the diagnosis model and the inference method. The *diagnosis model* represents the knowledge on how the identification of the fault cause is carried out. The elements of the model are causes and symptoms. The *inference method* is the algorithm that identifies the cause of the problems based on the value of the symptoms.

Defining the diagnosis model comprises two phases. Firstly, the qualitative model should be identified, that is, the causes and symptoms for diagnosis in a given technology (GSM, UMTS, multi-technology networks, etc.). Causes can be modelled as discrete random variables with two states: {absent / present}. Two types of symptoms are considered in the RAN of wireless networks: Alarms and KPIs. Alarms can also be modelled as discrete random variables with two states: {off / on}. KPIs are inherently continuous, but they can be modelled either as continuous or discrete random variables. In the latter case, the discretized KPI may have any discrete number of states, each representing a subset of the continuous range of the KPI, e.g. {normal / high / very high}.

Secondly, the quantitative model should be specified, that is, the parameters of the model. In the continuous model, probability density functions (pdfs) should be defined. If some predefined distributions are used, e.g. gaussian pdfs, the parameters of the model would be the parameters of those distributions, e.g. the average and the standard deviation. In the discrete model, the parameters are thresholds for the discretized KPIs and probabilities.

Once the quantitative and qualitative models have been defined, the inference method consists in calculating the probability of each possible cause. Given the value of the symptoms $E = \{S_1,...,S_M\}$, the probability of cause C_i, i=1…K, can be obtained as:

$$P(C_i|E)=\frac{P(C_i)\cdot\prod_{j=1}^{M} f(S_j|C_i)}{\sum_{k=1}^{K} P(C_k)\cdot\prod_{j=1}^{M} f(S_j|C_k)} \qquad (2)$$

where $P(C_i)$ are the prior probabilities of the causes and $f(S_j|C_i)$ are the probabilities of the symptoms given the causes. $f(S_j|C_i)$ are probability density functions in the continuous model and discrete probabilities in the discrete model. It can be noticed that the calculated probabilities, $P(C_i|E)$, are those required in eq.(1), $P(A_i = yes|E)$.

Eq.(2) has been obtained applying the Bayes' rule and taking into account the following assumptions: i) Only a cause can happen at the same time; ii) Symptoms are independent given the causes. These independence assumptions are realistic in the RAN of wireless networks, but even if they were not, this model has proven to give very good results (Rish, 2001). This model is called Naïve Bayes Model or Simple Bayes Model. More complex relations among the elements in the model are possible by using the theory of Bayesian Networks (BNs) (Jensen, 2001a; Pearl, 1988), also called Probabilistic Belief Networks. In (Barco, 2006), it was shown that for the RAN of wireless networks, the improvement in accuracy obtained when using a more complex model is negligible compared to the added complexity.

The most difficult task in model construction is the definition of the quantitative part. That implies defining the probabilities in eq.(2): $P(C_i)$ and $f(S_j|C_i)$ i=1…K, j=1…M. The prior probabilities of the causes, $P(C_i)$, can easily be elicited by experts or calculated from training data as the frequency of occurrence of each type of fault. Regarding the probability density functions, $f(S_j|C_i)$, both types of symptoms (KPIs and alarms) should be distinguished. Firstly, the pdf for an alarm given a cause is reduced to the probability of the alarm being active given the cause, which can also be obtained from training data by calculating the relative frequencies. Nevertheless, defining the conditional pdfs for the KPIs is much more complex because these symptoms are inherently continuous. Therefore, two alternative solutions have been proposed, which will be called continuous or discrete model, depending whether KPIs are modelled as continuous or discrete variables, respectively.

a. Continuous model.

In this model, KPIs are modelled as continuous variables. If the symptoms followed a known pdf, defining $f(S_j|C_i)$ would become a parameter estimation problem. Taking into account that the KPIs in the RAN are normally expressed as relative frequencies (e.g. "Percentage of samples in UL with RXLEV<10'', "Percentage of handovers due to bad quality''), their pdfs can be approximated by beta pdfs (Barco, 2005). Therefore, the only required information to completely define the conditional probabilities in the model are the parameters *a* and *b* of the beta pdfs for each pair symptom/cause. Those parameters can be obtained from training data by means of parametric estimation techniques, e.g. maximum likelihood estimation.

b. Discrete model.

In the discrete model, KPIs are discretized before being modelled as discrete symptoms. Thus, the parameters of this model are thresholds and probabilities.

On the one hand, the thresholds are interval boundaries for the discretization of the continuous symptoms. That is, $t_{j,k}$ is the k^{th} threshold for symptom S_j, which partitions it into states $s_{j,k}$ and $s_{j,k+1}$. Different methods have been proposed to define those thresholds:

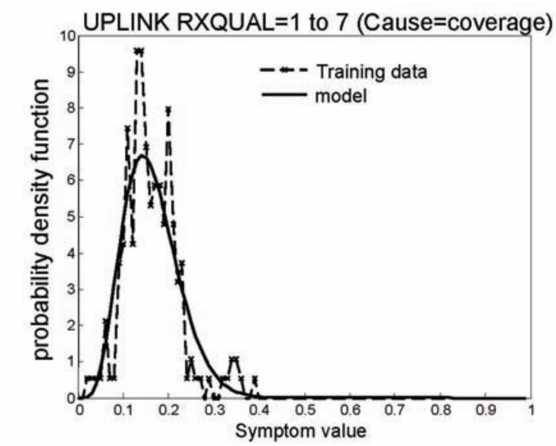
- Percentile-based discretization (PBD) (Khanafer, 2008): According to this method, a percentage is specified X%, e.g. 90%. Then, the X% percentile of the symptom values in a training set is the pursued threshold. The main advantage of this method is that it does not require including the real fault cause in the training set. However, the achieved diagnosis accuracy is lower than that obtained with the methods described as follows.
- Entropy Minimization discretization (EMD) (Fayyad, 1993): This method calculates the threshold for a given symptom by minimizing the entropy of the partition induced by that threshold.
- Selective Entropy Minimization discretization (SEMD) (Barco, 2008b): This method is a modification of EMD, which considers that each symptom is only related to some causes, meaning that in most cases only some causes lead to an anomalous value of the symptom.

On the other hand, once KPIs have been discretized, the pdfs $f(S_j|C_i)$ become discrete probabilities, $P(S_j|C_i)$. These probabilities can be obtained from training data by using different algorithms: Maximum Likelihood Estimation, Laplace's Law of succession (Khanafer, 2008) or M-estimate method (Barco, 2008b).

Study Cases

The auto-diagnosis system proposed in the previous section can be applied to any RAN of wireless networks, as long as the main causes and symptoms used for diagnosis are defined. In this section, some case studies performed on different networks will be described.

Figure 3. Approximation of symptom by beta pdf



a. In (Barco, 2005), an auto-diagnosis system for the RAN of GSM/GPRS networks was designed. Table 1 shows some causes and symptoms in the model. A continuous model using the principles explained in Section 3.4.a was built. Fig.3 shows the accurate approximation of a quality related symptom by a beta pdf. The system was tested in a live network, achieving a diagnosis accuracy of 71%.

b. In (Khanafer, 2008), an auto-diagnosis system was built for the RAN of UMTS networks. Table 2 shows some causes and symptoms in the model. A discrete model was defined according to the theory exposed in Section 3.4.b. Parameters of the model were obtained from training data using the following methods. The Laplace's Law of succession was used to calculate probabilities. In order to calculate the thresholds, two discretization methods were compared: PBD and EMD.

Testing the system was done in two steps. In the first stage, a network simulator was used as generator of test and training cases. In a second step, the auto-diagnosis system was tested in a live UMTS network. With the simulator, the diagnosis accuracy when using the PBD method was 59%, whereas the accuracy when using the EMD method was 70%. In the real UMTS network, the achieved diagnosis accuracy was 88% (using the PBD method).

c. In (Samhat, 2007), an auto-diagnosis system was defined for WLAN networks. Table 3 shows some causes and symptoms in the model. A discrete model was defined according to the theory exposed in Section 3.4.b. The discretization method used to obtain the thresholds was the PBD, whereas probabilities were obtained with a Maximum Likelihood Estimation. A simulator was used for generating training case from which specifying the diagnosis model.

Table 1. Examples of causes and symptoms in the RAN of a GSM/GPRS network

Causes	**Symptoms**
Interference in Uplink	Percentage of samples in UL with RXQUAL in bands 1 to 7 (BER>0.2%)
Interference in Downlink	Percentage of samples in UL with RXLEV<10 (level<-100 dBm)
Lack of coverage	Percentage of UL signal level handovers
Faulty TRX	Percentage of DL signal level handovers
Transcoder Fault	Percentage of UL quality handovers
Transmission Fault	Percentage of samples on idle channels out of interference band 1

Table 2. Examples of causes and symptoms in the RAN of a UMTS network

Causes	**Symptoms**
Channel element fault	Dropped Call Rate
Wrong Admission control parameters	Blocked Call Rate
Pilot Power too low	Macrodiversity Blocking Rate
Antenna tilt too high	Capacity
Wrong Mobility parameters	Active Set Update Rate
Wrong Congestion control parameters	Received Signal Strength

Table 3. Examples of causes and symptoms in a WLAN network

Causes	Symptoms
Bad coverage	Receive signal Strength Indicator
High noise	Noise level
High load	Number of users
Broken antenna	Throughput
MAC buffer problem	FCS errors

FUTURE TRENDS

Although network management has always played a key role for industry, only recently it has received a similar level of attention from the research community. In the context of self-managing networks, a framework for automated fault management in wireless networks has been presented in this chapter. The proposed systems are intended to improve operational efficiency in current and future networks, by means of the following benefits:

- Automated troubleshooting reduces the time required to identify the fault cause in cells with problems. This means that network performance is enhanced as the downtime and the time with reduced quality of service is significantly limited. Consequently, the number of unsatisfied users decreases and the threat posed by other cellular network operators and other emerging technologies is counteracted.
- Thanks to the automation of TS procedures, operational costs are reduced because fewer personnel with a lower skill level are required. This is due to the fact that the majority of problems can be rectified with the help of the automated tool. The knowledge of highly experienced staff, which is released from this work, can be utilized for other aspects of network optimization, thereby further increasing network performance.
- One of the main difficulties in research on auto-diagnosis is the lack of documentation about fault management in wireless networks. The steps followed to diagnose the fault cause in a cell are normally known only by a few experts who acquired that knowledge from hands-on experience. Hence, one additional benefit of the methods proposed in this chapter is that the TS knowledge can be stored in the auto-diagnosis tool, therefore making the company less vulnerable to staff fluctuations.

It is important to stress those engineering aspects learnt from working with cellular network operators, which may help anyone who attempts to create a similar system for wireless networks or any other application domain in which a database of cases is not available. The most important issues to be taken into account when building an auto-troubleshooting system are the following:

- A close cooperation between network operators, manufacturers of network equipment and academic researchers has proved to be essential for the success of the project. On the one hand, the contribution from network operators is the expertise in the application domain (i.e. troubleshooting in the RAN of wireless networks). On the other hand, the researchers may contribute with their knowledge on artificial intelligence and wireless networks. The manufacturers of network equipment provide their expertise on the counters generated by their equipment and they may be also responsible of implementing prototype tools.
- When creating a knowledge-based model, a feedback cycle is required in order to fine-tune the model parameters:
    - o Firstly, the artificial intelligence technique (e.g. probabilistic networks) should be selected.

- Then, the model parameters should be provided by the experts in the application domain (e.g. experts in fault management in the RAN) with the help of the knowledge engineers[4] or an automatic tool (Barco, 2009b). This stage, despite being conceptually simple, is one of the most difficult ones in model building. The reason is that TS experts are used to a language completely different to the terms utilised by the knowledge engineers. Consequently, it is recommended that knowledge engineers have also a background in the domain under study (mobile communications in this case), so that knowledge transfer can be simplified. In addition, the nature of the tasks performed by TS experts prevents them from spending sufficient time building a model. This is especially true if they do not see the immediate benefits of having an automatic system. In this phase, it is important to reduce the number of parameters that experts have to specify.
- Once the model is defined, it has to be tested in a cellular network. That implies that a prototype TS tool should be available at this stage. The model should be evaluated on some faulty cells and its results should be compared with the diagnosis supplied by an expert. If the diagnosis of the automatic tool is incorrect, then the TS expert together with the knowledge engineers should analyse which parameters of the model should be changed so that the expected results are obtained. This procedure should be repeated as many times as required. This can be a very time-consuming process. Thus, it requires a strong commitment from the operator so that experts do not give up after analysing a few cases.

In conclusion, apart from the theoretical contributions, the importance of this chapter lies in bringing together two completely separated worlds, such as theoretical research and operational network management. In order to deploy auto-diagnosis systems, it is essential to cooperate with several parties, each of them with its own interests and ways to solve problems.

The auto-troubleshooting system (TSS) can work independently from the NMS, but most of the benefits of auto-troubleshooting are achieved when the TSS is an integrated part of the NMS. This integrated solution will provide direct access to information required in fault analysis as well as access to the operators' fault management system. An integrated solution is also beneficial in case of TS of multi-vendor networks and of multi-system networks (GSM, UMTS, WLAN). The TS expert system is system independent. Thus, with minor modifications multi-vendor and multi-system networks can be supported. The integration in the NMS guarantees that the TSS is fully synchronised with the complete fault management system. Hence, all relevant TS cases can be automatically directed to the TSS so that if it finds the solution, the case is cleared, reported and filed. If the problem is not found by the expert system, it can be redirected to the specialists for further analysis and the final conclusions can be incorporated into the knowledge of the expert system.

Future lines in automated fault management will focus on defining diagnosis models for new radio access technologies. In addition, a big effort should be invested in performing trials in real networks in order to fine-tune models and methods. New methods may be studied in order to enhance diagnosis accuracy when the techniques presented in this chapter are sufficiently tested.

CONCLUSIONS

Automation of management processes has become a key strategy for the telecommunication business. In wireless networks, this subject has been driven by the increasing complexity of wireless technologies and operational tasks. In this scenario, this chapter has presented an automated fault management system for the RAN of wireless networks.

Some concepts related to fault management have been clarified and it has been described how fault management is performed in current cellular networks. These definitions and procedures were not documented in previous bibliography, despite the fact that troubleshooting is one of the main activities of operators of wireless networks.

The most complex task in fault management is diagnosis. This chapter has also presented an auto-diagnosis system based on probabilistic models. Some trials in real networks have proven the feasibility of the proposed techniques.

The described methods can be used to increase operational efficiency in current and future wireless networks (GSM, GPRS, UMTS, WLAN, B3G networks, ...). These techniques can also be extended to other telecommunication areas where automation of fault management is required.

REFERENCES

Altman, Z., Skehill, R., Barco, R., Moltsen, L., Brennan, R., Samhat, A., Khanafer, R., Dubreil, H., Barry, M., & Solana, B. (2006). The Celtic Gandalf Framework. *IEEE Mediterranean Electrotechnical Conference*, (pp. 595-598).

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S. (1987). MUNIN: A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the International Joint Conference on Artificial Intelligence*. Milan, Italy, (pp. 366-372).

Barco, R., Nielsen, L., Guerrero, R., Hylander, G., & Patel, S. (2002a). Automated troubleshooting of a mobile communication network using Bayesian networks. *Proceedings of the IEEE International Workshop on Mobile and Wireless Communications Networks (MWCN'02)*. Stockholm, Sweden (pp. 606-610).

Barco, R., Guerrero, R., Hylander, G., Nielsen, L., Partanen, M., & Patel, S. (2002b). Automated troubleshooting of mobile networks using Bayesian Networks. *Proceedings of the IASTED International Conference on Communication Systems and Networks (CSN'02)*. Málaga, Spain (pp. 105-110)

Barco, R., Wille, V., & Díez, L. (2005). System for automated diagnosis in cellular networks based on performance indicators. *European Transactions on Telecommunications*, *16*(5), (pp. 399-409).

Barco, R., Wille, V., Díez, L., & Lázaro P. (2006). Comparison of probabilistic models used for diagnosis in cellular networks. *Proceedings of the IEEE Vehicular Technology Conference VTC'06*. Melbourne, Australia.

Barco, R., Díez, L., Wille, V., & Lázaro, P. (2008a). Automatic diagnosis of mobile communication networks under imprecise parameters. *Expert Systems with Applications*, 36(1), 489-500, Jan.2009..

Barco, R., Lázaro, P., Díez, L., & Wille, V. (2008b). Continuous versus discrete model in auto-diagnosis systems for wireless networks. *IEEE Transactions on Mobile Computing*, 7(6), 673 - 681, Jun.2008.

Barco, R., Lázaro, P., Wille, V., Díez, L. and Patel, S. (2009b). Knowledge acquisition for diagnosis model in wireless networks. *Expert systems with Applications*. In press "

Barreto, G. A., Mota, J. C. M., Souza, L. G. M., Frota, R. A., Aguayo, L., Yamamoto, J. S., & Oliveira, P. (2004). Competitive Neural Networks for Fault Detection and Diagnosis in 3G Cellular Systems. *Lecture Notes in Computer Science (LNCS)*, *3124*, 207-213.

Barreto, G. A., Mota, J. C. M., Souza, L. G. M., Frota, R. A., & Aguayo, L., Lehtimäki, P., Hätönen, K., & Simula, O. (2005). Condition monitoring of 3G cellular networks through competitive neural models. *IEEE Transactions on Neural Networks*, *16*(5), 1064-1075.

Breese, J., & Heckerman, D. (1996). Topics in Decision-Theoretic Troubleshooting: Repair and Experiment. *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence* (pp. 124-132). Portland, Oregon.

Cayrac, D., Dubois, D., Haziza, M., & Prade, H. (1994). Possibility theory in fault mode effects analyses - A satellite fault diagnosis application. *Proceedings of the IEEE International Conference on Fuzzy Systems* (pp. 1176-1181). Orlando, Florida.

Deng, R. H., Lazar, A., & Wang, W. (1993). A Probabilistic Approach to Fault Diagnosis in Linear Lightware Networks. *IEEE Journal on Selected Areas in Communications*, *11*, 1438-1448.

Fayyad, M., & Irani, K.B. (1993). Multi-interval discretization of continuous valued attributes for classification learning. *Proceedings of the International Joint Conference on Artificial Intelligence* pp. 1022-1027). Chambery, France.

Halonen, T., Romero, J., & Melero, J. (2003). *GSM, GPRS & EDGE Performance Evolution Towards 3G/UMTS*. Chichester, England: Wiley.

Heckerman, D., Horvitz, E., & Nathwani, B. (1992). Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine*, 31, 90-105.

Heckerman, D., Breese, J., & Rommelse, K. (1995). Decision-theoretic troubleshooting. *Communication of the ACM*, *38*(3), 49-57.

Hoglund, A. J., Hatonen, K., & Sorvari A. S. (2000). A computer host-based user anomaly detection system using the self-organizing map. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, *5*, 411-416. Como, Italy.

Jakobson, G., & Weissman, M. D. (1993). Alarm Correlation. *IEEE Network* (pp. 52-59).

Jamalipour, A., Wada, T., & Yamazato, T. (2005). A tutorial on multiple access technologies for beyond 3G mobile networks. *IEEE Communications Magazine*, *43*(2), 110-117.

Jensen, F. (2001a). *Bayesian Networks and decision graphs*. New York, USA: Springer-Verlag.

Jensen, F. V., Kjærulff, U., Kristiansen, B., Langseth, H., Skaanning, C., Vomlel, J., & Vomlelová, M. (2001b). SACSO methodology for troubleshooting complex systems. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 5(4), 321-333.

Johnson, J. L. (2004). *Fundamentals of Cellular Network Planning and Optimisation*. West Sussex, England: Wiley.

Katzela, I., & Schwartz, M. (1995). Schemes for fault identification in communication networks. *IEEE/ACM Transactions on Networking*, *3*(6), 753-764.

Khanafer, R., Moltsen, L., Dubreil, H., Altman, Z., & Barco, R. (2006). A Bayesian approach for automated troubleshooting for UMTS networks. *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06)*. Karlsruhe, Germany.

Khanafer, R., Solana, B., Triola, J., Barco, R., Moltsen, L., Altman, Z., & Lázaro, P. (2008). Automated diagnosis for UMTS networks using Bayesian Network Approach. *IEEE Transactions on Vehicular Technology 57*(4), 2451-2461, Jul. 2008..

Kephart, J., & Chess, D. (2003). The vision of autonomic computing. *Computer, 36*(1), (pp. 41-50).

Laiho J., Wacker, A., & Novosad, T. (2002a). *Radio network planning and optimisation for UMTS*. Chichester, UK: Wiley

Laiho, J., Kylväjä, V., & Höglund, A. (2002). Utilisation of Advanced Analysis Methods in UMTS Networks. *Proceedings of the IEEE Vehicular Technology Conference VTC'02* (pp. 726-730). Birmingham, USA

Laiho, J., Ravio, K., Lehtimäki, P., Hätönen, K., & Simula, O. (2005). Advanced Analysis Methods for 3G Cellular Networks. *IEEE Transactions on Wireless Communications, 4*(3), 930-942.

Lázaro, P., Barco, R., & Hermoso, J. M. (2002) Diagnosis of Earth Stations Using Bayesian Networks. *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*(pp. 268-272). Málaga, Spain.

Lehtimäki, P., & Raivio, K. (2005a). A SOM Based Approach for Visualization of GSM Network Performance Data. *Proceedings of the International Symposium on Intelligent Data Analysis.* Madrid, Spain.

Lehtimäki, P., & Raivio, K. (2005b). A Knowledge-Based Model for Analyzing GSM Network Performance. *Proceedings of the International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*. Bari, Italy.

Lempiäinen, J., & Manninen, M. (2001). *Radio interface system planning for GSM/GPRS/UMTS*. Kluwer Academic Publishers. Dordrecht, The Netherlands.

Madruga E. L., & Tarouco L. M. R. (1994). Fault management tools for a cooperative and decentralized network operations environment. *IEEE Journal on Selected Areas in Communications, 12*(6), 1121-1130.

Melchiors, C., & Tarouco L. M. R. (2000). Troubleshooting network faults using past experience. *Network Operations and Management Symposium*. Honolulu, Hawaii (pp. 549-562).

Ng, G., & Ong, K. (2000). Using a qualitative probabilistic network to explain diagnostic reasoning in an expert system for chest pain diagnosis. *Computers in Cardiology* (pp. 569-572).

Pang, B., Zhang, D., Li., N., & Kuanquan, W. (2004). Computerized tongue diagnosis based on Bayesian networks. *IEEE Transactions on Biomedical Engineering, 51*(10), 1803-1810.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, California: Morgan Kaufmann.

Pras, A., Schowalder, J., Burgess, M., Festor, O., Pérez, G. M., Stadler, R., & Stiller, B. (2007). Key research challenges in network management. *IEEE Communications Magazine, 45*(10), 104-110.

Rish, I. (2001). An empirical study of the naive Bayes classifier. *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 41-46). Seattle, USA.

Samhat, A., Skehill, R., & Altman, Z. (2007). Automated Troubleshooting in WLAN Networks. *Proceedings 16th IST Mobile and Wireless Communication Summit.* Budapest, Hungary.

Skaanning, C., Jensen, F. V., & Kjaerulff, U. (2000b). Printer troubleshooting using Bayesian Networks. *Lecture Notes in Artificial Intelligence, 1821*, 367-379.

Skaanning, C. (2000a). A Knowledge Acquisition Tool for Bayesian-Network Troubleshooters. *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence* (pp. 549-557). Stanford, USA.

Steinder, M., & Sethi, A. S. (2004). Probabilistic Fault Localization in Communication Systems Using Belief Networks. *IEEE/ACM Transactions on Networking, 12*(5), 809-822.

Strassner, J. (2004). Autonomic networking – Theory and practice. *IEEE/IFIP Network Operations and Management Symposium*.

Wietgrefe, H., Tuchs, K., Jobmann, K., Carls, G., Fröhlich P., Nedil, W., & Steinfeld, S. (1997). Using Neural Networks for Alarm Correlation in Cellular Phone Networks. *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications* (pp. 7.2.97.1-10). Melbourne, Australia.

Wietgrefe, H. (2002). Investigation and practical assessment of alarm correlation methods for the use in GSM access networks. *Proceedings of the IEEE/IFIP Network Operations and Management Symposium* (pp. 391- 403). Florence, Italy.

Yemini, S. A., Kliger, S., Mozes, E., Yemini, Y., & Ohsie, D. (1996). High speed and robust event correlation. *IEEE Communications Magazine*, 34, 82-90.

KEY TERMS

Alarm Correlation: Filtering of alarms to meaningful high level alarms in order to avoid overloading the operators. It involves different tasks: reduction of multiple occurrences of an alarm into a single alarm, inhibition of low priority alarms in the presence of higher priority alarms, substitution of a specific set of correlated alarms by a new one, etc.

Auto-Diagnosis: Automated identification of the fault cause that is causing problems, based on the analysis of the symptom values.

Bayesian Network: A directed graph and a set of conditional probability functions that allow an efficient representation of a joint distribution over a set of random variables.

Decision-Theoretic Troubleshooting: Method to solve problems that tries to find a sequence of tests and actions that maximizes the efficiency of the troubleshooting process.

Diagnosis Accuracy: Percentage of cases in a data set correctly classified, that is, the percentage of cases where the diagnosed cause is equal to the actual cause.

Diagnosis Model: It is a representation of how the identification of the fault cause should be performed. It comprises a qualitative part (causes and symptoms) and a quantitative part (parameters that quantify the relations among causes and symptoms).

Fault Detection: Identification of those cells in a cellular network that have some problems.

Key Performance Indicators: Quantifiable measurements that reflect the performance of the wireless network.

Troubleshooting: Also referred as fault management. Procedures carried out to solve problems. It comprises three phases: fault detection, diagnosis and solution deployment.

Trouble Ticket: Procedure used to reflect a problem status, when a fault is investigated. It consists of the fault description and the steps performed so far to solve it out or the identification of the faulty equipment in case the problem is deemed as a hardware fault.

ENDNOTES

1. The Radio Access Network (RAN) is the part of the network in charge of providing access to the users and connecting them to the core network. The wireless access is one of the key characteristics of mobile communication networks, which identifies them and differentiates them from traditional wired networks.
2. The diagnosis accuracy is the percentage of cases in a data set correctly classified, that is,

the percentage of cases where the diagnosed cause is equal to the actual cause.

[3] See (Heckerman, 1995) for detailed description of assumptions and demonstration for optimal repair sequence.

[4] The knowledge engineer is a person, but could also be a computer program, that interprets the information presented by experts in the application domain and defines a model based on that information.

Chapter XLVIII
Modeling Intrusion Detection with Self Similar Traffic in Enterprise Networks

Cajetan M. Akujuobi
Prairie View A&M University, USA

Nana K. Ampah
Prairie View A&M University, USA

ABSTRACT

Most of the existing networks (e.g., telecommunications, industrial control, enterprise networks etc.) have been globally connected to open computer networks (Internet) in order to decentralize planning, management and controls in business. Most of these networks were originally designed without security considerations, thereby making them vulnerable to cyber attacks. This has given rise to the need for efficient and scalable intrusion detection systems (IDSs) and intrusion prevention systems (IPSs) to secure existing networks. Existing IDSs and IPSs have five major limitations, which prevent them from securing networks absolutely. It has been proven that the right combination of security techniques always protects networks better. This approach used change in Hurst parameter and a signal processing application of wavelets (i.e., multi-resolution technique) to develop an IDS. The novelty of our proposed IDS technique presented in this chapter lies in its efficiency and ability to eliminate most of the limitations of existing IDSs and IPSs, thereby ensuring high level network protection.

INTRODUCTION

Telecommunications networks form the major part or the foundation of all business enterprise networks, which may include a combination of local area networks (LANs), metropolitan area networks (MANs), wide area networks (WANs), and remote LAN access connectivity. Business enterprise networks are the main targets for hackers due to the fact that most financial transactions (i.e., E-commence) take place online and the networks also handle vast amounts of data and other resources (Satti & Garner, 2001). Handling transactions online is on the increase everyday because it makes life easier for both the customers as well as the enterprises offering services (Tront & Marchany, 2004). Business enterprise networks also have lots of bandwidth, which is very attractive to hackers because they take advantage of that by using those networks as launching pads to attack others (Tront & Marchany, 2004; Janakiraman, et. al., 2003). It therefore becomes very difficult for the IDSs and IPSs at the receiving end to detect and prevent hackers, since the packet header information will indicate legitimate senders. This is the main reason why most IPSs are easily bypassed by hackers (Tront & Marchany, 2004). Intrusion prevention, which is a proactive technique, prevents attacks from entering the network. Unfortunately, some of the attacks still bypass the IPSs. Intrusion detection, on the other hand, detects attacks only after they have entered the network.

The increasing use of Internet for various economic activities coupled with the complex and dynamic nature of network security management has given rise to numerous attacks on the network itself as well as any other networks connected to it. There is also a rapid increase in the daily use of data networks for research and development collaborations with respect to rapidly changing technologies. Securing information on data networks has therefore become a very difficult task considering the diverse types and number of intrusions being recorded daily. The situation has necessitated drastic research work in the area of network security, especially in the development of IDSs and IPSs intended to detect and prevent all possible attacks on a given network. The development of IDSs and IPSs has therefore acquired increasing commercial importance (Janakiraman, et. al., 2003; Akujuobi & Ampah, 2007; Akujuobi, et. al., 2007). Although attacks are generally assumed to emanate from outside a given network, the most dangerous attacks actually emanate from the network itself. Those are really difficult to detect, since most users of the network are assumed to be trusted people. There is no existing security technique that guarantees total security for a given network, so the best approach frequently used is to implement several layers of techniques.

As a second line of defense, a combination of IDS techniques is required to back-up the existing IPSs. This has been a difficult task for network administrators mainly due to the availability of different types of IDSs on the market. These IDSs use either anomaly-based or signature-based detection techniques. Anomaly detection techniques detect both known and unknown attacks, but signature-based detection techniques detect only known attacks. The main approaches of anomaly detection techniques are statistical, predictive pattern generation, neural networks, and sequence matching and learning. The main approaches of signature-based detection techniques are expert systems, keystroke monitoring, model-based, state transition analysis, and pattern matching (Biermann, et. al., 2001). This chapter also investigates the negative effects of designing, planning and managing telecommunication networks, industrial control networks, and business enterprise networks with special emphases on issues like effectiveness, efficiency and reliability without considering proper security planning, management and constraints.

These networks have become vulnerable due to their recent connectivity to open networks with the intention of establishing decentralized management and remote control (Chunmei, et. al., 2004; Chi-Ho & Kwong, 2005; Amanullah, et. al., 2005). Cyber attacks on control systems for power, water, oil/gas, chemical, paper and agriculture businesses recorded in the past included denial of service attacks. Some confirmed cyber attacks included intentionally opening valves resulting in discharge of millions of liters of sewage, opening breaker switches, tampering with boiler control settings resulting in shutdown of utility boilers, shutdown of combustion turbine power plants, and shutdown of industrial facilities (Amanullah, et. al., 2005).

This chapter makes references to the practical examples of such control systems like SCADA (Supervisory Control and Data Acquisition) networked systems used in most corporate enterprises. Security threats to such networks fall under two major categories as follows:

- **External threats:** Unauthorized remote connections to gateway or servers from Internet or Public Switched Telephone Network (PSTN), thereby gaining process control privileges through misconfigured access control policies.
- **Internal threats:** Buffer overflow or program flaw attacks from disgruntled workers to control servers, thereby gaining root privileges (Chi-Ho & Kwong, 2005).

The objective of this chapter is to introduce an efficient and scalable IDS/IPS for business enterprise networks, which will guarantee a very high level of protection from attacks. In order to achieve the aforementioned objective, the proposed security technique must satisfy the following:

- Detect and prevent Distributed Denial of Service (DDoS) attacks based on SYN-Flood attacks;
- Detect and prevent SYN-Flood attacks (i. e. one of the most dangerous Denial of Service (DoS) attacks, which has no countermeasure); and
- Eliminate four out of the five limitations of existing IDSs.

It is important to note that DDoS attacks based on SYN-Flood attacks is one of the most dangerous network attacks, which can shut down a whole network no matter its size.

BACKGROUND

The following major approaches are used to manage network security problems as stated earlier:

- Intrusion Detection (Traditional); and
- Intrusion Prevention (Proactive).

The basic techniques used by the two approaches are as follows:

- Signature based detection system (Attack patterns are considered as signatures);
- Anomaly detection system (Anything unusual is considered as suspect);
- Distributed intrusion detection system (Data is collected and analyzed in a distributed fashion);
- Centralized intrusion detection system (Data is collected in a distributed fashion but analyzed centrally).

There is no existing IDS or IPS that can detect or prevent all intrusions. For example, configuring a firewall to be 100% foolproof compromises the very service provided by the network. The use of conventional encryption algorithms and system level security techniques have helped to some extent, but not to the levels expected (Fadia, 2006). The novel IDS technique presented in this chapter attempted to eliminate most of the following five limitations associated with existing IDSs (Satti & Garner, 2001):

- **Use of central analyzer:** Whenever the central analyzer is attacked by an intruder the whole system will be without protection, so it becomes a single point of failure (Janakiraman, et. al., 2003);
- **Limited scalability:** Processing all data at a central point limits the size of the entire network that can be monitored and controlled at a time. Data collection in a distributed fashion also causes excessive traffic in the network;

- **Effectiveness:** The ability of existing IDSs/IPSs to detect and prevent intrusion is still not clearly established because of high false positive and false negative rates (Chunmei, et. al., 2004);
- **Efficiency:** Quantifying resources like time, power, bandwidth, and storage used by existing IDSs will be a critical success factor (Khoshgoftaar & Abushadi, 2004); and
- **Security:** Securing the security data itself from intruders is also a very important limitation to existing IDSs.

This work was motivated by the pressing need to develop a network security system, which will eliminate most of the aforementioned limitations leading to a very high level of protection from attacks, thereby reducing losses incurred by all kinds of business enterprises, and also to help homeland security in their global security efforts. Most research work on security management systems can be classified as IDS only technique, IPS only technique and a combined IDS/IPS technique. Our approach is based on the combined IDS/IPS class. Examples of the IDS only approach can be found in (Ramana, et. al., 2007). Examples of the IPS only approach can be found in (Bruschi, et. al., 2007) and examples of the combined IDS/IPS approach can be found in (Sher & Magedanz, 2007).

MAIN FOCUS OF THE CHAPTER

This technique was based on network packet behavior leading to network-based intrusion detection. It employed anomaly detection as its analysis strategy. The self-similarity property of real network traffic was used together with the signal detection abilities of wavelets in detecting attacks. This technique also attempted to reduce the effectiveness of distributed attacks, which deny authorized users access to system resources. Securing of all network security data, which is an important limitation to existing IDSs and IPSs was also ensured by this technique.

Two scenarios were considered: conventional approach and new approach. The scenario for the conventional network-based anomaly (IDS); employed change in Hurst parameter as its parameter for detection just like the new approach. This was because the goal also focused on the application of multi-resolution techniques and its effects on IDSs and IPSs. This approach employed multi-resolution techniques, which formed the basis for the first scenario. The general goal of this approach was to design and develop a network-based intrusion detection system, which used anomaly detection techniques. Unlike all anomaly detection techniques, which depend on just the behavior of the network, this system employed a quantitative approach. The quantity used was the change in Hurst parameter of the network traffic at specific points in the network after an attack or intrusion (Nash & Ragsdale, 2001). No matter the type of attack traffic present, there will always be a change in Hurst parameter registered, thereby enabling this IDS to detect the attack. The novelty of this approach lies in the fact that no existing IDS employs the change in Hurst parameter as a parameter for detecting intrusion. Multi-resolution technique was used to transmit network security data from the network to the Central Detection Point instead of transmission through the network itself (Nash & Ragsdale, 2001; Garcia, et. al., 2002). The transmission can also be effected by any means (i. e., wired or wireless) at the implementation stage.

Multi-resolution technique was used to decompose, transmit and reconstruct signals from the enterprise network to a Central Detection Point for further analysis. The outcome of all new detections to be made by this technique will eventually be used to update the attack list of the operating IPSs (Graps, 1995). The first scenario applied multi-resolution techniques, but the second did not. Fig. 1 depicts the implementation scheme for this new technique. It differs from the conventional technique because there is only one link between the Central Router Device and the Central Detection Point in that case, whereas there are several separate links between each IDS and the Central detection point as depicted in Figure 1. Only two links were shown for the sake of clarity.

Self-Similar Stochastic Process

The m-aggregated time series

$X^{(m)} = \{X_k^{(m)}, k = 0, 1, 2,...\}$ is defined as:

$$X_k^{(m)} = \frac{1}{m}\sum_{i=km-(m-1)}^{km} X_i \qquad (1)$$

If the process has the same statistical properties at all values of m (all aggregations), then that process is self-similar. Self-similarity for a process is defined in terms of its var[X(t)] and autocorrelation $R(t_1, t_2)$:

$$Var[X(t)]=E[X^2(t)]-\mu^2(t), \qquad (2)$$

where μ - The mean of X.

$$R(t_1,t_2)=E[X(t_1)X(t_2)] \qquad (3)$$

A process X is exactly self-similar with parameter β (0 < β < 1) if for all m = 1, 2, …,

$$Var(X^{(m)}) = \frac{Var(X)}{m^\beta} \qquad (4)$$

$$R_{X^{(m)}}(X) = R(t_1, t_1 + k) = R_X(k) \qquad (5)$$

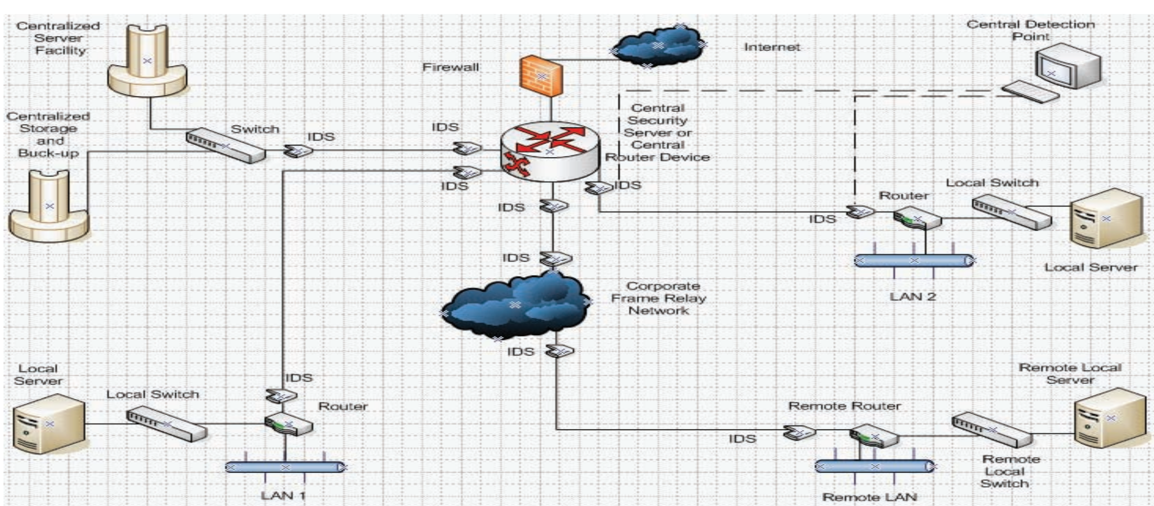
In many cases a weaker definition is needed: A process X is asymptotically self-similar with parameter β (0 < β < 1) if for all k large enough,

$$Var(X^{(m)}) = \frac{Var(X)}{m^\beta} \qquad (6)$$

$$R_{X^{(m)}}(k) \to R_X(k) \text{ as } m \to \infty \qquad (7)$$

The variance of a self-similar process decreases proportional to $1/m^\beta$ as m approaches infinity. Equation 7 shows that the autocorrelation of the aggregated process has the same form as the original one, which suggests that the degree

Figure 1. New approach with network-based IDS.



of variability is the same at all time resolutions. The variable

$$H = 1 - \frac{\beta}{2}, \quad 0 < \beta < 1 \tag{8}$$

is known as the Hurst parameter, and gives the degree of self-similarity of a process. When H = 0.5, self-similarity does not exist. The degree of self-similarity increases as H approaches one. Network traffic has been proven to exhibit self-similar properties (Shibin, et. al., 2004).

Modeling Technique

A reasonable model for estimating the change in Hurst parameter, ΔH is based on the model for detecting a DC signal or level in the presence of White Gaussian Noise, WGN (Kay, 1993). That model is represented by the following equation:

$$x[n] = A + w[n],\ n = 0, 1, ..., N - 1 \tag{9}$$

where

$x[n]$ - Data for the DC signal or level with WGN (with intrusion);
A - The DC signal level to be estimated;
$w[n]$ - WGN samples with each sample having the PDF $N(0, \sigma^2)$, which denotes a Gaussian distribution with a zero mean and a variance of σ^2. Please note that intrusion is represented by White Gaussian Noise under this model.

The expression for the change in Hurst parameter with and without noise will be:

$$\Delta H = H_1 - H \tag{10}$$

where ΔH - Change in Hurst parameter, which is assumed to be always positive;
H_1 - Hurst parameter of signal with WGN (intrusion);
H - Hurst parameter of signal without WGN (no intrusion).

Re-writing equation (10) in the form of equation (9) gives:

$$H[n] = \Delta H + H_1[n],\ n = 0, 1, ..., N - 1 \tag{11}$$

Note that there is no negative sign in front of ΔH, since it is assumed to be always positive. Note also that, in determining the PDF of H[n], it is obvious that during a SYN-Flood attack or a distributed denial of service attack based on SYN-Flood attacks, the PDF of the distribution of H_1[n] is dominated by the PDF of noise (or intrusion). This implies that although the H_1[n] distribution is made up of noise and H[n], noise (or intrusion) dominates during a SYN-Flood attack or a distributed denial of service attack based on SYN-Flood attacks. The rest of the modeling will therefore not include the H[n] factor in H_1[n] on the right hand side of equation (11).

The probability density function of the distribution $x[n]$ will be:

$$p(x;A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n] - A)^2\right]$$

$$p(x;A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A^2)\right] \tag{12}$$

Similarly, the PDF of the distribution, H[n] will be:

$$p(H;\Delta H) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(H[n] - \Delta H^2)\right] \tag{13}$$

The log-likelihood function of equation (13) will be:

$$\ln p(H;\Delta H) = -\ln[(2\pi\sigma^2)^{N/2}] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(H[n]-\Delta H)^2$$

$$\ln p(H;\Delta H) = -\frac{N}{2}\ln[(2\pi\sigma^2)] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(H[n]-\Delta H)^2 \quad (14)$$

The maximum likelihood estimation (MLE) of ΔH is found by equating the derivative of the log-likelihood function (i.e. equation (14)) to zero:

$$\frac{\partial \ln p(H;\Delta H)}{\partial \Delta H} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(H[n]-\Delta H) = 0 \quad (15)$$

$$\sum_{n=0}^{N-1}(H[n]-\Delta H) = 0$$

$$\sum_{n=0}^{N-1}H[n] - \sum_{n=0}^{N-1}\Delta H = 0$$

$$\sum_{n=0}^{N-1}H[n] - N\Delta H = 0$$

$$\therefore \Delta\hat{H} = \frac{1}{N}\sum_{n=0}^{N-1}H[n] \quad (16)$$

Although the maximum likelihood procedure yields an estimator that is asymptotically efficient, it also sometimes yields an efficient estimator for finite data records. It follows from equation (16) that the estimation of ΔH is the same as finding the sample mean of $H[n]$, which is already known to be an efficient estimator. Hence the MLE of ΔH found here is efficient. For the purpose of this work, values of ΔH is used instead of $H[n]$ in order to analyze the effect of ΔH on detection. It is also clear from equation (11) that, H[n] is directly proportional to ΔH for any given value of "n".

Equation (15) can be re-written as follows:

$$\frac{\partial \ln p(H;\Delta H)}{\partial \Delta H} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(H[n]-\Delta H) = \frac{N}{\sigma^2}(\bar{H}-\Delta H) \quad (17)$$

where

$\bar{H}$ - The sample mean of $H[n]$.

To find the Crame-Rao Lower Bound (CRLB) of the estimator ΔH, we need to take the second derivative of equation (17):

$$\frac{\partial^2 \ln p(H;\Delta H)}{\partial \Delta H^2} = -\frac{N}{\sigma^2} \quad (18)$$

The variance of any unbiased estimator $\hat{\theta}$ must satisfy the following expression:

$$Var(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]} \quad (19)$$

where the derivative is evaluated at the true value of θ and the estimation is taken with respect to p(x;θ). But,

$$-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right] = -\left[-\frac{N}{\sigma^2}\right] = \frac{N}{\sigma^2} \quad (20)$$

since N/σ^2 is a constant. From expression (19) and equation (20), we get the following:

$$Var(\Delta\hat{H}) \geq \frac{\sigma^2}{N} \quad (21)$$

An unbiased estimator may be found that attains the bound for all θ if and only if:

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = I(\theta)(g(x)-\theta) \quad (22)$$

for some functions g and I. That estimator, which is the minimum variance unbiased estimator (MVUE) is $\hat{\theta} = g(x)$, and the minimum variance is $1/I(\theta)$, where $I(\theta)$ is the Fisher Information. Comparing equations (17) and (22), it is clear that the MVUE of ΔH (i. e., $\Delta\hat{H}$) is the sample mean of $H[n]$(i. e., $\overline{H}$) just as obtained for the MLE case and $I(\theta)$ will be equal to

$$\frac{N}{\sigma^2} \text{ or } \frac{1}{Var(\Delta\hat{H})}.$$

It is clear from the analysis made in finding the MLE and CRLB for $\Delta\hat{H}$ that the proposed estimator, $\Delta\hat{H}$ is asymptotically unbiased and asymptotically achieves the CRLB. Hence, it is asymptotically efficient.

Simulation Studies

New Approach (with Multi-resolution Techniques)

In this approach, MATLAB codes for Pareto Distribution was used to generate the network traffic, but real traffic will be used at the implementation stage of this technique. Fig. 2 shows the graph of the generated network traffic with a target Hurst Parameter, H_T of 0.75. MATLAB codes were used to simulate this technique.

The Rescaled Adjusted Range Statistic Analysis (i.e. R/S Analysis) was used to test the generated self-similar input traffic. It estimated the Hurst Parameter (i. e., Calculated H_E) and compared it to H_T. Fig. 3 shows the results after calculating the estimated Hurst Parameter, H_E.

Note that the gradient of the best straight line was the same as H_E. The equation of the best straight line from the simulation was:

$$Y = 0.6848X - 0.9196$$

so, $H_E = 0.6848$

As stated earlier, the final analysis of the network traffic was carried out at the Central Detection Point. Transmission of traffic from the

Figure 2. Self-similar input traffic or input network traffic without intrusion (for H_T =0.75)

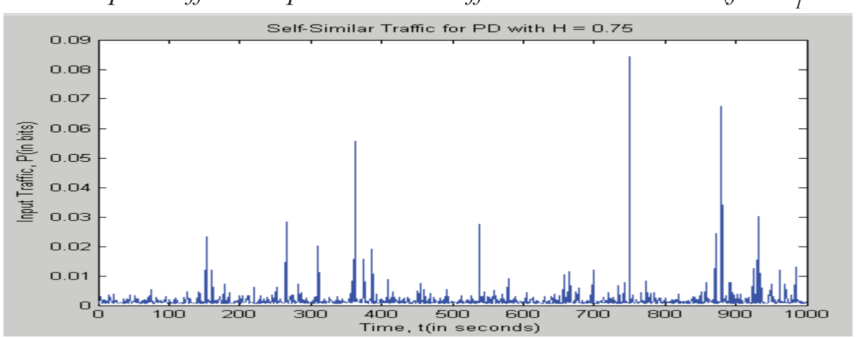


Figure 3. Best straight line obtained from testing the generated traffic without intrusion or noise

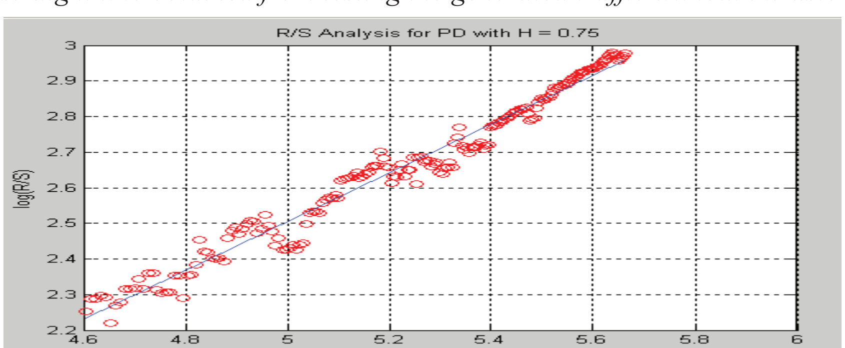
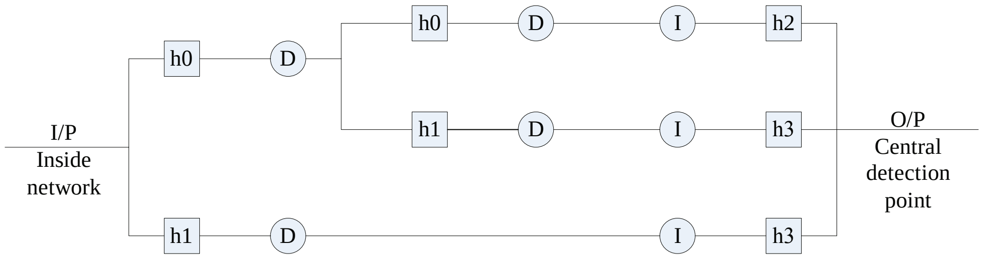


Figure 4. Scheme for the multi-resolution technique



network to the Central Detection Point was done using a one-dimensional, two-stage multi-resolution technique. Haar wavelets were applied here. Fig. 4 shows the scheme for the multi-resolution technique used.

The Rescaled Adjusted Range Statistic Analysis (i.e. R/S Analysis) was again used to test the received self-similar output traffic. The estimated Hurst Parameter, H_E was calculated for the received output signal and compared to that obtained during the test before the transmission.

The gradient of the best straight line was the same as $H_{E.}$ The equation of the best straight line from the simulation was:

$$Y = 0.3704X + 1.1204$$

so, $H_E = 0.3704$

The value for H_E should have been closer to its value for the test before the multi-resolution technique (i. e., H_E =0.6848) as observed for the other values of H_T. But, since Pareto Distribution itself involves the use of random number generation, it is expected to occasionally experience such an anomaly. The worst network attacks are known to be distributive by nature, so all the generators considered in modeling intrusion generate distributive noise. Gaussian Noise Generator was chosen to model intrusion because unlike Rayleigh Noise Generator and Rician Noise Generator, it generates distributive noise with given mean and variance values. The others generate distributive noise without given mean and variance values, hence they are less flexible to use (Garcia, et. al., 2002). The chosen Gaussian Noise Generator added Additive White Gaussian Noise (AWGN) to the input signal with the assumption that the signal-to-noise ratio was 10dB.

Testing (Here, H_E was not compared to any targeted value as before.) and transmission of the generated self-similar input traffic with intrusion or noise was done as before. Testing of the received self-similar output traffic with intrusion or noise was also done as before. (But here, the estimated Hurst Parameter, H_E was compared to that obtained during the test before the transmission.) Also, the gradient of the best straight line (i. e. for the generated signal with intrusion or noise) was the same as $H_{E.}$ The equation of the best straight line from the simulation was:

$$Y = 0.6349X - 0.5914$$

so, $H_E = 0.6349$

The gradient of the best straight line (i. e. for the output signal after applying multi-resolution techniques) was the same as $H_{E.}$ The equation of the best straight line from the simulation was:

$$Y = 0.6255X - 0.2321$$

so, $H_E = 0.6255$

Everything done so far was for $H_T = 0.75$. An extension from this stage compared H_E values before and after transmission for input traffic without intrusion (noise) and for input traffic with intrusion (noise). The values of H_T used for this analysis were: 0.55, 0.65, 0.75, 0.85, and 0.95. Table 1 shows the results of all the tests and the changes in Hurst parameter for all H_T. For a better analysis of the results, ten sets of values were used instead of one. This helped to substantiate the aforementioned changes involving the Hurst Parameters of the output signals without and with intrusion. The Hurst Parameters of output signals without and with intrusion in Figure 5 are the most important statistics out of the lot investigated because they clearly showed differences or changes in the various values.

For the purpose of this work, arithmetic mean and standard deviation were applied to verify how the changes were related to each other across 0.55 < H < 0.95. Mean and standard deviation were calculated for each of the 10 simulations. Table 2 shows three sets of results indicating the changes in Hurst parameter for the chosen H_T values out of the ten simulations. The mean and standard deviation values were calculated using the following equations:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{23}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{24}$$

Table 1. Results of all tests and changes in Hurst parameter

Targeted Hurst parameter	Signal without intrusion (noise)		Signal with intrusion (noise)		Change in Hurst parameter (outputs)
	Input Signal	Output Signal	Input Signal	Output Signal	$\Delta H_{E4}=H_{E2}-H_{E4}$
H_T	H_{E1}	H_{E2}	H_{E3}	H_{E4}	
0.55	0.54	0.45	0.64	0.46	-0.01
0.65	0.40	0.52	0.49	0.67	-0.16
0.75	0.68	0.37	0.63	0.63	-0.26
0.85	0.60	0.49	0.53	0.51	-0.01
0.95	0.46	0.47	0.60	0.50	-0.04

Figure 5. Hurst Parameter for O/P signals without/with intrusion.

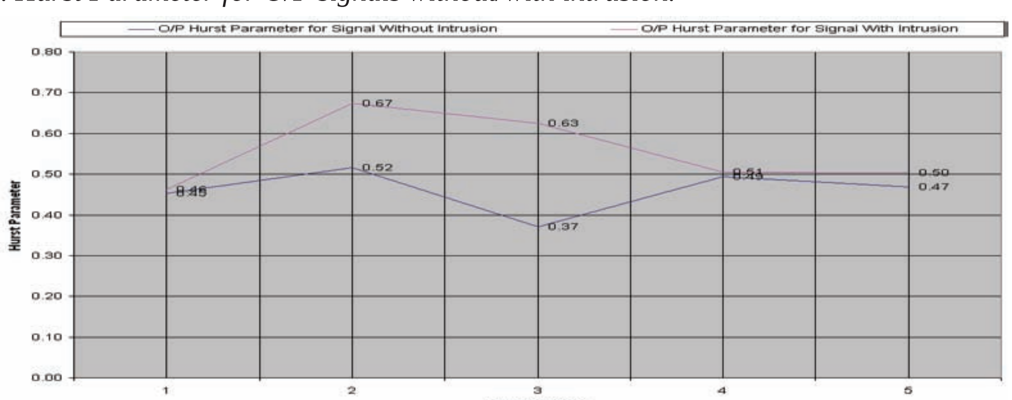


Table 2. Results from three out of ten tests showing changes in Hurst parameter, mean and standard deviation

H_T	H_{E2}	H_{E4}	$\Delta H_{E4}=H_{E2}-H_{E4}$		Mean	Standard deviation
			Raw value	Absolute value		
First set of data						
0.55	0.45	0.46	-0.01	0.01	0.1	0.1
0.65	0.52	0.67	-0.16	0.16		
0.75	0.37	0.63	-0.26	0.26		
0.85	0.49	0.51	-0.01	0.01		
0.95	0.47	0.50	-0.04	0.04		
Second set of data						
0.55	0.60	0.75	-0.15	0.15	0.16	0.1
0.65	0.60	0.29	0.31	0.31		
0.75	0.44	0.47	-0.03	0.03		
0.85	0.48	0.39	0.09	0.09		
0.95	0.41	0.63	-0.22	0.22		
Third set of data						
0.55	0.66	0.64	0.02	0.02	0.11	0.07
0.65	0.50	0.40	0.10	0.1		
0.75	0.45	0.37	0.08	0.08		
0.85	0.52	0.63	-0.11	0.11		
0.95	0.61	0.37	0.24	0.24		

The nominal value of change in Hurst Parameter was the minimum value that indicated the presence of intrusion. This value was the mean of all the means (i. e., for the 10 sets of simulations) calculated from Table 3. The standard deviation needed was calculated for the 10 sets of mean to ensure that all its values were really close to the new mean or nominal value of change (Graps, 1995). Although there were mean values lower than the mean of all the ten mean values, the target here was to detect the most serious distributive attacks, which will definitely cause large changes in Hurst parameter.

From Table 3, the new mean or the nominal value of change, ΔH and the standard deviation, σ_x were found as follows:

Table 3. Mean and standard deviation for ten simulations

Simulation number	Mean	Standard deviation
1	0.096	0.0989
2	0.16	0.098
3	0.11	0.0721
4	0.07	0.039
5	0.174	0.0587
6	0.136	0.106
7	0.062	0.0461
8	0.092	0.0736
9	0.076	0.0485
10	0.106	0.0618

$\Delta H = 0.108$ and $\sigma_x = 0.0358$

Considering the aforementioned value for ΔH and the range of "H" used in the simulation studies (i. e., 0.55 < H < 0.95), the nominal value of change in percentage was calculated as follows:

$$\Delta H(\%) = \frac{0.108}{(0.95-0.55)} \times 100\% = 27\%$$

Conventional Method (Without Multi-Resolution Techniques)

Everything done before this stage involved the application of multi-resolution techniques in decomposing, transmitting and reconstructing signals from the network to the Central Detection Point. The whole exercise was repeated, but without using multi-resolution techniques. In this case, the input signals were the same as the output signals because of the absence of the multi-resolution techniques. This implies that all the previous codes were reused up to and not beyond the input stages. The nominal value of change in Hurst Parameter was found just as in the first scenario. The new mean or the nominal value of change, ΔH and the standard deviation, σ_x were found as follows:

ΔH = 0.1084 and σ_x = 0.0302

Considering the aforementioned value for ΔH and the range of "H" used in the simulation studies (i. e., 0.55 < H < 0.95), the nominal value of change in percentage was calculated as follows:

$$\Delta \text{H}(\%) = \frac{0.1084}{(0.95 - 0.55)} \times 100\% = 27.1\%$$

Discussions

The standard deviations for all the 10 sets of values (with and without multi-resolution techniques) were almost the same and very close to zero. This means that the changes or differences in Hurst Parameter were very close to each other across all values of "H" (i. e., 0.55 < H < 0.95). Also, the nominal values of change in Hurst Parameter, ΔH (with and without multi-resolution techniques) were equal to 0.108 and 0.1084 respectively. The corresponding standard deviations, σ_x were also 0.0358 and 0.0302 respectively. This also proved that both 10 sets of mean values were close to each other and also close to ΔH, since both σ_x values were close to zero.

The $\Delta H(\%)$ values (with and without multi-resolution techniques) which were equal to 27% and 27.1% respectively of the range of "H" used for the entire simulation, represented effective minimum changes beyond which an intrusion can be detected.

Therefore, setting a minimum value of $\Delta H(\%)$ ensured effective detection at the Central Detection Point making $\Delta H(\%)$ a strong parameter for intrusion detection. This will further help reduce high false positive and false negative rates, leading to the elimination of limitation number three (i. e. effectiveness). Finally, $\Delta H(\%)$ values for both scenarios (with and without multi-resolution techniques) were the same due to the unique merit of multi-resolution techniques in signal analysis.

FUTURE TRENDS

The following areas will be considered for further investigation:

- Implementation of technique (i. e., Where to place detectors in the network.);
- To consider both the pdfs of H and noise for H_1 in determining the PDF of H[n];
- The effect of different signal-to-noise ratios after adding AWGN to the input traffic on the change in Hurst parameter and the intrusion detection criterion;
- The use of confidence interval in the estimation of change in Hurst parameter;
- The best type of multi-resolution technique to use (i. e., 1-D, 1-Stage or 1-D, 2-Stage or 1-D, 3-Stage etc.);
- The best type of wavelet to use (i. e., Haar, Daubechies, Morlet etc.);
- Using of real network traffic and real intrusion for better results;
- Power studies for output signals without and with intrusion;
- Choice of noise generator to be reconsidered for improvement if possible;
- Choice of network traffic generator to be reconsidered for improvement if possible; and
- Other statistics should be considered apart from mean and standard deviation.

CONCLUSION

In conclusion, the telecommunication network security issues discussed in this chapter are very important to the emerging telecommunications security management for business enterprises. The proposed approach is better than previous techniques as discussed earlier in this chapter for the following reasons. Nash and Ragsdale (2001) used Hurst parameter, but they did not investigate how the presence of intrusion affected Hurst parameter and further helped in detecting intrusion. The technique used by Garcia, et. al. (2002) depended on TCP/IP connections, which can be compromised by an intruder launching attacks from a trusted host (zombie).

The scenario with multi-resolution techniques has the following advantages. The analysis of data after detection can be done anywhere (i.e., way beyond the network). Transmission of security data after detection can be done separately without going through the network itself thereby saving bandwidth leading to the elimination of part of limitation number four (i.e., efficiency). Handling security data separately from the main network makes it more difficult for attackers to attack the security network or system itself leading to the elimination of limitation number five (i.e., security), and partly eliminating limitation number two (i.e., limited scalability: reduction of traffic in actual network).

REFERENCES

Akujuobi, C. M., & Ampah, N. K. (2007). Enterprise network intrusion detection and prevention system (ENIDPS), *Society of Photographic Instrumentation Engineers Defense and Security Symposium*, *6538*, Orlando, Florida.

Akujuobi, C. M., Ampah, N. K., & Sadiku, M. N. O. (2007). An Intrusion Detection Technique Based on Change in Hurst Parameter with Application to Network Security, *International Journal of Computer Science and Network Security*, *7, I(5),* 55–64.

Amanullah, M. T. O., Kalam, A., & Zayegh, H. (2005). Network security vulnerabilities in SCADA and EMS, *Transmission and Distribution Conference and Exhibition: Asia and Pasific*, (pp. 1-6). Dalian.

Biermann, E., Cloete, E., & Venter, L. M. (2001). A comparison of intrusion detection systems. *Journal on Computers and Security, 20, I(8)*, 676-683.

Bruschi, D., Cavallaro, L., & Lanzi, A. (2007). An effective technique for preventing Mimicry and Impossible Paths Execution Attacks, *IEEE International Conference on Performance, Computing, and Communications,* (pp. 418-425). New Orleans, LA: Dipt. di Informatica e Comunicazione, Univ. degli Studi di Milano, Milan.

Chi-Ho, T., & Kwong, S. (2005). Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. *IEEE International Conference on Industrial Technology,* (pp. 51-56). Department of Computer Science, City University of Hong Kong, China.

Chunmei, Y., Mingchu, L., Jianbo, M., & Jizhou, S. (2004). Honeypot and scan detection in intrusion detection system, in *Proceedings of Canadian Conference on Electrical and Computer Engineering*, *2,* 1107–1110. School of Electronic Information Engineering, Tianjin University, China.

Fadia, A. (2). (2006). *Network Security: A Hacker's Perspective*. Boston, MA: Thomson Course Technology.

Garcia, R. C., Sadiku, M. N. O., & Cannady, J. D. (2002). WAID: wavelet analysis intrusion detection. 45th. Midwest Symposium on *Circuits and Systems, 3,* 688–691. Science Atlanta, GA, USA.

Graps, A. (1995). An introduction to wavelets. *Computer Science and Engineering, 2, I(2),* 50-61. 22724 Majestic Oak Way, Cupertino, CA, USA.

Janakiraman, R., Waldvogel, M., & Qi Z. (2003). Indra:a peer-to-peer approach to network intrusion detection and prevention, *Twelfth IEEE International Workshops on Enabling Technologies:*

Infrastructure for Collaborative Enterprises, (pp. 226 -231). Dept. of Comput. Sci. & Eng., Washington Univ., St. Louis, WA, USA.

Kay, S. M. (1). (1993). *Fundamentals of Statistical Signal Processing*: *Estimation Theory*, Prentice Hall Signal Processing Series (pp. 1-217).

Khoshgoftaar, T. M. & Abushadi, M. E. (2004). Resource-sensitive intrusion detection models for network traffic. *Eighth IEEE International Symposium on High Assurance Systems Engineering*, (pp. 249 – 258). Department of Computer Science & Engineering, Florida Atlantic University, Boca Raton, FL, USA.

Nash, D. A., & Ragsdale D. J. (2001). Simulation of self-similarity in network utilization patterns as a precursor to automated testing of intrusion detection systems. *IEEE Transactions on Systems, Man and Cybernetics, 31,* I(4), 327–331. Dept. of Electr. Eng. & Comput. Sci., US Mil. Acad., West Point, NY, USA.

Pollet, J. (2002). Developing a solid SCADA security strategy. *2nd. ISA/IEEE Sensors for Industry Conference*, 2002, (pp. 148-156). PlantData Technologies.

Ramana, R. K., Singh, S., & Varghese, G. (2007). On scalable attack detection in the network", *IEEE/ACM Transactions on Networking, 15,* I(1), 14-25. Department of Computer Science, University of California San Diego, La Jolla, CA.

Satti, M. M., & Garner, B. J. (2001). Information security on internet enterprise managed intrusion detection system (EMIDS). *IEEE International Multi-Topic Conference,* (pp. 234 -238). Sydney, NSW, Australia: Macquarie Corporate Telecommunication.

Sher, M., & Magedanz, T. (2007). Protecting IP Multimedia Subsystem (IMS) server delivery platform from Time Independent Attacks. *3rd International Symposium on Information Assurance and Security,* (pp. 171-176). Manchester, UK: Technical University, Berlin, Germany.

Shibin S., Ng, J. K.-Y., & Bihai, T. (2004). Some results on the self-similarity property in communication networks, *IEEE Transactions on Communications, 52,* I(10), 1636-1642. Dept. of Risk Manage. & Insurance, Zhongshan Univ., Guangzhou, China.

Tront, J. G., & Marchany, R. C. (2004). Internet Security: Intrusion Detection and Prevention. *37th Annual Hawaii International Conference on System Sciences,* (pp. 188-188). Bradley Dept. of Electrical & Computer Engineering, Virginia Tech., Blacksburg, VA.

KEY TERMS

E-Commerce: The buying and selling of products or services over electronic systems such as the Internet and other computer networks.

Enterprise Network: A large corporate network which spans multiple sites nationwide and possibly worldwide including LANs, MANs, and WANs depending on the specific needs of a given enterprise.

Intrusion Detection System: A computer security system that detects unwanted manipulations of computer systems, mainly through the Internet.

Intrusion Prevention System: A computer security system that monitors network and/or system activities for malicious or unwanted behavior and can react, in real-time, to block or prevent those activities.

Local Area Networks: A computer network covering a small geographic area, like a home, office, or group of buildings e.g. a school.

Metropolitan Area Networks: A network that connects two or more local area networks together but does not extend beyond the boundaries of the immediate town, city, or metropolitan area.

Wide Area Networks: A data communications network that covers a relatively broad geographic area (i.e. one city to another and one country to another country) and that often uses transmission facilities provided by common carriers, such as telephone companies.

Chapter XLIX
High-Speed Multimedia Networks: Critical Issues and Trends

Dimitris Kanellopoulos
University of Patras, Greece

ABSTRACT

This chapter presents high-speed networking technologies and standards such as Asynchronous Transfer Mode (ATM), Fast Ethernet, 10 Gigabit Ethernet, Synchronous Optical Network (SONET), Resilient Packet Ring (RPR), Provider Backbone Transport (PBT), Provider Backbone Bridges (PBB), Transport - Multi Protocol Label Switching (T-MPLS) and Optical Transport Network (OTN). It considers the requirements imposed to high-speed networks by multimedia applications and analyses crucial issues of high-speed networking such as bandwidth problems, discarding policies and fast broadcast. Finally, the chapter discusses future trends in high-speed multimedia networking.

INTRODUCTION

High-speed networks are speeding up computer communications and are designed to transmit continuous-media traffic such as audio and video (Mammeri and Lorenz, 2004). Continuous media and data can be integrated in a multimedia service. In the past decade, an integrated communication fabric and spectacular increases in bandwidth have been achieved in order to support multimedia applications. For example, Chen et al. (1999) developed a high-speed network environment to satisfy the special requirements of multiparty multimedia applications such as real-time communication, multicast transmission and media synchronization. High-speed multimedia networks carry continuous-media traffic and pose many design challenges not faced in other sorts of telecommunication systems. Nowadays, it is possible to carry efficiently multimedia over high-speed networks because these networks support the quality of service (QoS) requirements imposed by multimedia applications (Gibson, 2000). These QoS requirements are specified by the following four closely related parameters: 1) bandwidth on demand, 2) low end-to-end delay, 3) low delay variation (or delay jitter) and 4) acceptable error or loss rate without retransmission as the delay would be unacceptable with retransmission. These parameters may all be required in a multimedia application.

This chapter will present high-speed networking technologies and will analyze crucial issues of high-speed networking such as bandwidth problems, discarding policies and fast broadcast. In addition, it will present various categories of multimedia applications and will discuss future trends in high-speed multimedia networking.

BACKGROUND

In the 1990s, many multimedia services were developed on a single network infrastructure, while many QoS mechanisms emerged in order to support real-time and interactive applications (Aurrecoechea et al., 1998). Giordano et al. (2003) describe the current evolution of QoS architectures, mechanisms and protocols in the Internet, as it is ongoing in the framework of the European Union funded research projects (AQUILA, CADENUS, TEQUILA) on premium Internet Protocol (IP) networks. Recently, an IP based global information infrastructure (GII) was established, which was increasingly based on fast packet switching technology interconnected by fibber optic cable. The IP joined previously disjoint networks, while the World Wide Web (Web) became the killer application that drove bandwidth demand. Today, customers access not only data in Web pages but images and streaming multimedia content using Web browsers.

High-speed networking research moved up the protocol stack to be more concerned with multimedia applications. In the late 1990s, we had the practical application of fast packet switching technology to IP routers, which became IP switches. This was caused by the failure of Asynchronous Transfer Mode (ATM) and the decreasing cost in hardware. This divergence of high-speed networking research into the application layer and switch design had the effect of fragmenting the discipline into other sub-disciplines of communications such as router/switch design and multimedia applications. Two forces resisted the global deployment of a connection-oriented network layer such as ATM.

- In the mid 1990s, the explosion of the Internet and Web entrenched TCP as the end-to-end protocol and IP as the single global network layer.
- The limitations of shared medium link protocols such as Ethernet and token ring were overcome by the evolution of Ethernet to a switched point-to-point link protocol with order-of-magnitude increases in data rate. This evolution additionally reduced the motivation for adoption of ATM using scalable Synchronous Optical Network (SONET) links to increase the bandwidth on network links.

The important characteristics of fast packet switching technologies began to be incorporated into the Internet. For example, IP switches based on the fast switch fabrics and protocol optimisations such as Multi Protocol Label Switching (MPLS) began to be established (Armitage, 2000).

HIGH SPEED NETWORKS

Pillalamarri and Ghosh (2005) write… *"A high-speed network is one where its link bandwidth, the source traffic controls that guarantee QoS for all traffic sources, and the total processing times within the nodes are such that the transmission time of a characteristic frame's transportation across its characteristic distance, is always less than the physical propagation delay of electro-magnetic transmission for that distance in that* *medium, irrespective of where the measurement is obtained in the network".*

High-speed networking requires high bandwidth and low latency. In addition, it requires the ability to manage with product paths of high bandwidth and low delay (Sterbenz and Touch, 2001). Inter-application delay forces the requirement for high bandwidth, low latency networks. This attitude is the motivation for why we need high-speed networking. End-to-end latency must be very low and bandwidth must be very high in order the transmission delay of data to be bounded. The design of a high-speed network requires:

- The use of an optical transmission medium and components.
- Long propagation delay compared to the transmission delay.
- Low error rate.
- Limited processing and abundant bandwidth.

The architecture and protocols of high-speed networks have the following features:

- *Simple algorithms.* The network protocols should be simplified as much as possible in order to reduce processing overhead.
- *End-to-end protocols.* We have the shift of processing from the internal nodes of the network to the edge nodes.
- *Regular network topologies* are preferred in order to simplify routing and thus to reduce network processing.
- *Limiting buffers.* It is impractical to provide many buffers to store large number of packets that will flow through the transmission links.
- *Simplified error control.* The error rate of a fibre-optical system is very low.
- *Trade-offs between bandwidth and delay.*

ASYNCHRONOUS TRANSFER MODE (ATM)

ATM is essentially a hybrid of packet-switched and circuit-switched network technologies (Sexton and Reid, 1997). ATM is similar to packet switching in that the flow of information is divided into streams of fixed-size cells, which are multiplexed and switched in the network. However, routing for cells in the same connection uses the same path. A cell consists of a 48-bytes information field and a 5-bytes header field. User payload is carried on the information field. The header field is used to provide control information such as the identification of the cell, the properties of the user payload (the priority and the length of user payload on the information field), error monitoring on the header field, routing and switching information etc.

ATM networks show high efficiency in broadband network access as they use statistical multiplexing dynamically to distribute the network resources among a wide spectrum of users. Statistical multiplexing is more efficient than other multiplexing methods because merging traffic reduces burstiness. ATM uses the virtual path (VP) concept to simplify the routing protocol and associated network architecture and to reduce the size of routing tables in the transit nodes. Furthermore, ATM allows the bandwidth to be dynamically allocated to each virtual connection (VC) based on the requirements of multimedia applications. In particular, various classes of traffic with different bit-rates and different QoS requirements can be handled in an ATM network that provides a set of traffic control capabilities such as network resource management, connection admission control, usage parameter control or policing function, network parameter control, priority control, traffic shaping, fast resource management, and congestion control.

FAST ETHERNET

Fast Ethernet was developed by the same standards body that produced the original Ethernet (IEEE 802.3). The Fast Ethernet additions are a supplement to the IEEE 802.3 standard, referred to as IEEE 802.3u (IEEE, 1995). To reflect its 100 Mbps transmission rate, Fast Ethernet is denoted 100BASE-T. In Fast Ethernet, the physical size of the collision domains is constrained to be approximately an order of magnitude smaller than the one of conventional Ethernet. Fast Ethernet holds up switched-hub operation, which can be used to get better performance from a multimedia viewpoint. In a shared media hub, a single 100 Mbps transmission channel is shared among all connected stations so that only one transmission can proceed at a time. This is accomplished by copying an incoming transmission on any port to all other ports (in order that it can be sensed by the CSMA/CD algorithm).

10 GIGABIT ETHERNET

Gigabit Ethernet (GigE) is a data link and physical layer technology. The IEEE 802.3z standard was completed in 1998 and it is a 1 Gbps, backward compatible, extension to the IEEE 802.3 standards for 10 and 100 Mbps Ethernet. GigE uses the same frame format as its 10 and 100 Mbps predecessors with frames of 64–1518 bytes excluding preamble and start-of-frame delimiter (SFD) and a 96 bit inter-frame gap (IFG). It represents another 10-fold increase in data rate and, again, reduces the time required to transmit a frame by a factor of 10. This could have been achieved by an additional reduction in network diameter. However, since a 20m network diameter was considered impractical, the IEEE 802.3z working committee, basically, redefined the media access control (MAC) layer for GigE by adding a mechanism to make a 200m network diameter possible at 1 Gbps. This mechanism is known as "*carrier extension*" (IEEE 802.3, 1998).

Whenever a shared gigabit network adapter transmits a frame shorter than 512 bytes long, it adds a new carrier extension field of up to 448 bytes (3584 bits) to the frame and continues to monitor for collisions while sending this special signal. The carrier extension field follows the cyclic redundancy check (CRC) field and contains a sequence of special non-data "extended carrier" symbols that are not considered part of the frame. The CRC remainder is calculated only on the original frame (i.e. without extension symbols) and the frame plus carrier extension lasts for a minimum of 512 bytes. To avoid late collisions for the required distance limits, GigE also extends the Ethernet slot time to 512 bytes (4096 bits), from 64 bytes (512 bits) for Ethernet and Fast Ethernet. The 64-byte minimum frame length and 96 bit IFG have not changed for GigE and frames longer than 512 bytes are not extended. Figure 1 shows the Gigabit Ethernet frame format when Carrier Extension is used.

10 Gigabit Ethernet (or 10GbE or 10 GigE) is the most recent and fastest of the Ethernet standards (Cheng et al., 2005). It defines a version of Ethernet with a nominal data rate of 10 Gbps, ten times as fast as Gigabit Ethernet. 10GbE over fiber and InfiniBand "like" copper cabling are specified by the IEEE 802.3-2005 standard. 10GbE over twisted pair has been released under the IEEE 802.3an amendment. 10 Gigabit Ethernet abandons half duplex links and repeaters (and

Figure 1. Format of gigabit Ethernet frame with carrier extension.

8	6	6	2	46-1500	4	0-448
Preamble + SFD	Destination address	Source address	Type/ Length	Data + Pad	CRC	Carrier Extension

<---------------------------- 512 byte minimum ------------------------------->

the CSMA/CD that goes with them) in favor of a system of purely full duplex links connected by switches as was already the normal practice with gigabit Ethernet. Kirstädter et al. (2006) focus on functionality and standards required to enable carrier-grade Ethernet-based core networks and discuss possible Ethernet backbone network architectures.

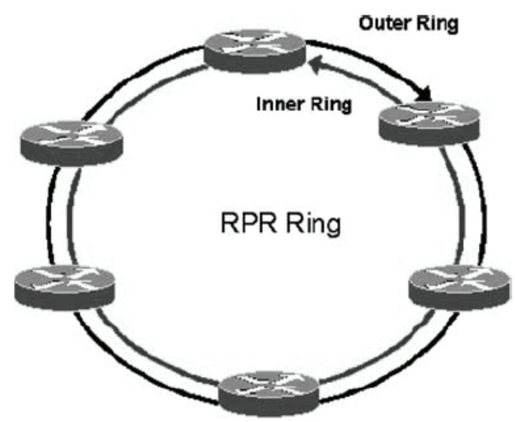
SYNCHRONOUS OPTICAL NETWORK (SONET)

Currently most telecom carriers use SONET/synchronous digital hierarchy (SDH) equipment to "aggregate" data traffic in the metropolitan area network (MAN) before accessing the Internet backbone network. SONET defines signal formats in optical networks (Sexton and Reid, 1997). The basic features include a frame signal format, signal rates which are multiples of 51-84 Mbps, and synchronous multiplexing. SONET provides networking services on the physical layer that defines the signal format, multiplexing/demultiplexing hierarchy, the schemes for performance monitoring, and fault tolerance of signal and network elements. The basic signal in the SONET hierarchy is the synchronous transport signal-level 1 (STS-1) which has a bit rate of 51-84 Mbps. SONET is a physical layer standard used for the data transmission in WAN environments. It supports data transmission in fiber optics with data rates in the range 51.8 Mbps until 2.5 Gbps. SONET is frequently the physical layer of B-ISDN (Broadband ISDN) and its data link layer is LAPD (Link Access Protocol - Channel D), while the time slice for a frame of SONET is 125 μsec. A similar technology to SONET is SDH.

RESILIENT PACKET RING (RPR)

Ring is the main topology for deploying high-speed MANs. The main reason is the inherent resiliency and fault tolerance advantage of ring networks. Specifically, by employing bidirectional connections in a ring, all nodes can remain connected even in the presence of a link failure. Resilient packet ring (RPR) is a ring-based architecture consisting of two optical rotating rings: one is referred to as the *inner ringlet*, and the other the *outer ringlet* (Figure 2).

Figure 2. The resilient packet ring



RPR standardized as IEEE 802.17 is a new MAC protocol (IEEE 802.17, 2004) for metro-ring networks and supports three service classes:

- *High priority Class A* with guaranteed rate and jitter, which consists of two subclasses: A0 and A1.
- *Medium priority Class B* that is also divided into two subclasses: B-CIR (committed information rate) and B-EIR (excess information rate).
- *Low priority Class C* (the best error traffic).

In RPR, packets are removed from the ring at the destination so that different segments of the ring can be used simultaneously for different flows (Davik et al., 2004). Therefore, the *spatial reuse* (concurrent transfers over the same ring) feature is achieved. Enabling the spatial reuse feature introduces the challenge of guaranteeing fairness among the nodes sharing the same link. RPR supports spatial reuse which increases the achieved throughput but it can also result in congestion and starvation of nodes on the ring. For that reason, it is vital to utilize mechanisms to enforce a fair allocation of the ring bandwidth in RPR. In order to preserve fairness among nodes, a fairness algorithm is deployed at each RPR node. When a node detects congestion, it calculates a fair rate, which is advertised to all upstream nodes contributing to congestion. Upon receiving the fair rate, the upstream nodes limit the rate of their injected traffic to the advertised fair rate. As a result, the congested node can utilize the unused capacity and add its local traffic to the ring. In RPR, bandwidth is reserved for subclass A0. Subclasses A1 and B-CIR have allocated bandwidth which can be reclaimed if not used. However, subclass B-EIR and class C are subject to the fairness algorithm and called *fairness eligible (FE) traffic*.

PROVIDER BACKBONE TRANSPORT (PBT)

Provider Backbone Transport (PBT) is a set of enhancements to Ethernet technology that allows the use of Ethernet as a carrier class transport network (Allan et al., 2006). PBT enables the creation of connection-oriented Ethernet tunnels that allow service providers to offer dedicated Ethernet links with guaranteed, deterministic performance levels. PBT forwards traffic on full MAC+Virtual LAN (60 bit) address. VLAN tag is no longer network global, that is scaling issues are removed, while a range of VLANs can be used for bridging and another range for PBT.

PBT is designed to meet or exceed the functionality of MPLS RSVP (Resource Control and Reservation Protocol)-TE tunnels, but it disables the concept of flooding/broadcasting and spanning tree protocol (STP). The STP and MAC learning function are disabled and the forwarding is based on the static forwarding database (FDB) entries made via management commands. All broadcast packets are dropped. All Destination Lookup Failure (DLF) packets are dropped rather than broadcasted. PBT simplifies the operational administration and maintenance (OAM) by using additional extensions based on IEEE 802.1ag. It also provides extensions in order to provide path protection levels. Path protection is provided by using one work and one protect VLAN ID (VID). In case of work path failure, the source node swaps the VID value to redirect the traffic onto the preconfigured protection path within 50 ms. Conclusively, the key features of PBT are: 1) traffic engineering and resiliency, 2) secure, deterministic delivery, 3) service scalability, 4) operational simplicity, 5) Ethernet tunneling with full MPLS interoperability, 6) service and transport layer independence (i.e. the services inside the tunnel could be Ethernet, IP, MPLS pseudo-wires or VPLS).

PROVIDER BACKBONE BRIDGES (PBB)

This standard (IEEE 802.1ah) defines an architecture and bridge protocols compatible and interoperable with Provider Bridged Network protocols and equipment allowing interconnection of multiple Provider Bridged Networks, to allow scaling to at least 220 Service VLANs, and to support management including SNMP (Simple Network Management Protocol). PBB will complete the future work identified by P802.1ad, by providing a specific means for interconnecting Provider Bridged Networks. It will enable a Service Provider to scale the number of Service VLANs in a Provider Network by interconnecting the Service Virtual LANs, and provide for interoperability and consistent standards based management. This project is intended to facilitate the scaling of Provider Bridged P802.1ad networks using existing Bridged and Virtual Bridged Local Area

Figure 3. T-MPLS frame structure

Network technologies. Despite user demand and initial deployment of LAN-based backbones for connecting P802.1ad networks, there is currently no interoperability between different vendors or a coherent management framework for different techniques. Most major carriers (who will be the users of this standard) are currently deploying LAN-based service networks that need to be scaled to meet the demands both of transition from existing leased line service and expansion of multipoint services.

TRANSPORT MULTI PROTOCOL LABEL SWITCHING (T-MPLS)

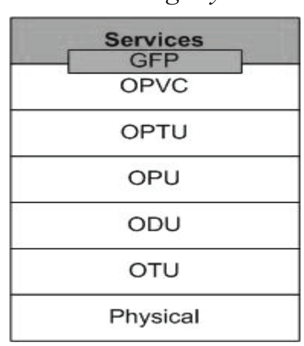
Transport - Multi-Protocol Label Switching (T-MPLS) is based on a profile of several existing standards, as MPLS already has many of the mechanisms in place to provide traffic engineering capabilities within a resilient architecture. T-MPLS was developed in response to the inherent complexity of MPLS implementations (ITU-T, 2005). By simplifying MPLS, T-MPLS provides the key functions necessary for traffic engineering and offers many advantages for carriers and service providers who are already proficient with MPLS technology. T-MPLS is an adaptation of MPLS and its main idea is to use the well-established MPLS concept known from IP routing and adapt it for transport forwarding issues (Lometti, 2006). In T-MPLS, pre-defined tunnels are established and an additional MPLS header is pushed in front of the client traffic that is transported transparently inside the backbone network. Similarly to Virtual LAN Cross-connect (VLAN-XC), the 20bit label is used to encode the backbone tunnel and is removed at the egress backbone switch. Figure 3 depicts the frame structure of T-MPLS.

OPTICAL TRANSPORT NETWORK (OTN)

Optical Transport Network (OTN) or digital wrapper technology is the ITU standard G.709 that was unified from competing standards being developed in both ITU and ANSI. An OTN is composed of a set of Optical Network Elements connected by optical fibre links, able to provide functionality of transport, multiplexing, routing, management, supervision and survivability of optical channels carrying client signals (Ciena, 2005). An OTN network consists of several networking layers (Figure 4).

The "Services" layer represents the end-user services such as GbE, SONET, SDH, FC (Fibre Channel) or any other protocol. For asynchronous services such as GbE or FC, the service is passed through a Generic Framing Procedure (GFP) mapper. The Optical channel Payload Virtual Container (OPVC) handles mapping the service into a uniform format. The OPVC is the only layer that needs to change to support a new service type. The Optical channel Payload Tributary Unit (OPTU) maps the output of the OPVC into a timeslot and performs timing adaptations to unify the clocking. The Optical channel Pay-

Figure 4. OTN networking layers



load Unit (OPU) contains all of the timeslots in the OTN frame. The Optical channel Data Unit (ODU) provides the path-level transport functions of the OPU. The Optical Transport Unit (OTU) provides the section-level overhead for the ODU and provides the GCC0 bytes. The Physical layer maps the OTU into a wavelength or wavelength-division multiplexing system. OTN is currently offered in three rates: OTU1, OTU2 and OTU3.

- OTU1 has a line rate of approximately 2.7 Gbps and was designed to transport a SONET OC-48 or synchronous digital hierarchy (SDH) STM-16 signal.
- OTU2 has a line rate of approximately 10.7 Gbps and was designed to transport an OC-192, STM-64 or 10Gbit/s WAN. OTU2 can be over-clocked (non standard) to carry signals faster than STM64/OC192 (9.953Gbit/s) like 10 gigabit Ethernet LAN PHY coming from IP/Ethernet switches and routers at full line rate (10.3 Gbps).
- OTU3 has line rate of approximately 43 Gbps and was designed to transport an OC-768 or STM256 signal.

MULTIMEDIA APPLICATIONS

Different multimedia applications have different QoS requirements and are classified into the following three categories (Lu, 2000):

- *Two-way conversational applications* including telephone and videophone services. This category is characterized by its stringent requirement on end-to-end delay which includes total time taken to capture, digitize, encode/compress audio/video data, transport them from the source to the destination, decode and display them to the user.
- Broadcasting services where the source is live. The main difference from the conversational applications is that it is one-way communication and it can tolerate more delay.
- Retrieval or on-demand applications where the user requests some stored items and the server delivers them to the user.

The characteristics of these application types should be used in designing and implementing respective applications in order to provide required QoS, while using network and system resources as efficient as possible. The user can specify the degree (or level) of guarantees. In general, there are three levels of guarantees:

- *Hard guarantee*: User-specified QoS should be met absolutely (100%). Reserving network and system resources based on the peak-bit rate of a stream achieves hard guarantees.
- *Soft or statistical guarantee*: User-specified QoS should be met to a certain specified percentage. This is suitable for continuous media, as these media normally do not need 100% accuracy in playback. This type of guarantee uses system resources more efficiently.
- *Best effort:* No guarantee is provided and the application is executed with whatever resources are available.

Additionally multimedia applications are categorised as follows:

- *Information access applications*, which are client/server applications with highly asymmetric bandwidth requirements and a fairly large granularity of transfers (e.g. Web and video on demand-VoD).
- *Multimedia applications that match the network*. If multimedia applications are aware of the underlying network infrastructure and control mechanisms, significant performance gains are possible. Matching applications with the network can dramati-

cally reduce the control overhead and data transformation.

- *Telepresence applications* involve the exchange of information that allows users to maintain a distributed virtual presence, frequently in the form of multimedia. Therefore, telepresence tends to be more symmetric in its bandwidth requirements and individual transfers are either at small granularity or a continuous stream.
- *Distributed computing applications* involve the distribution of computations beyond a room and involve an arbitrary exchange of data. While requirements are highly variable on the particular computation (which is in turn designed on network capabilities) in the general case the bandwidth, latency and synchronisation requirements can be very challenging.
- *Real-time electronic mail.* Some e-mail applications contain audio and low-resolution video as well as text. However, current e-mail services suffer from unpredictable delay owing to the available bandwidth in the network. Real-time e-mail supports video and audio conferencing.
- *Virtual reality applications* can be thought as dramatically, improved simulation. The idea is to simulate an environment so realistically that the user believes that (s)he is in the new environment. Virtual reality applications require high-speed networks for the following reasons: 1) the amount of data required to drive a visualization system is often large; 2) the output devices for the simulation environment may not all be attached to the same multimedia computer, but may have to be coordinated or even driven via the network; 3) several computers may have to exchange a lot of data to keep track of each participant's behaviour in the common area.
- *Cooperative design and writing.* In these applications, large amounts of images and data in addition to a voice signal need to be integrated and transmitted together with high-speed and reliability for effective collaboration with the cooperative members. Chen et al. (1999) have implemented a real-time conference system and a cooperative drawing tool (called CoXfig) which can be used by conferees to share a common canvas and produce complex figures.
- Other more complex application scenarios are compositions of the three core categories. For example, distance learning is a composition of telepresence and information access.

CURRENT ISSUES

In high-speed networks, switching techniques use the bandwidth available in the channel most efficiently. High performance switching is achieved by hardware developments (VLSI techniques and fibre optics) and by concepts such as distributed control, parallelism, self-routing etc. Oie et al. (1991) classified the switches into the following categories: 1) electronic versus photonic switching; 2) multiplexing techniques; 3) packet versus circuit switching; 4) synchronization techniques; 5) blocking versus non-blocking; 6) unicast versus multicast; 7) hardware versus software switching; 8) single-stage versus multi-stage switching; and 9) single-path or multi-path switching. As referred previously, the network and end-systems must provide a path of low-latency and high-bandwidth between multimedia applications to support low inter-application delay. However, real-world constraints make it difficult to provide high-performance paths to multimedia applications. These constraints include the speed of light, limits on channel capacity and switching rate, heterogeneity, policy and administration, cost and feasibility, backward compatibility, and standards.

BANDWIDTH PROBLEMS

In high-speed networks there are mainly four bandwidth problems. The first problem is topology design and bandwidth allocation. This problem is concerned with the ability to dynamically reconfigure a network in order to efficiently benefit from the network resources. The second problem is concerned with flow control and congestion avoidance. Bandwidth management protocols are used to prevent congestion, essentially by accepting or refusing a new packet arrival. The third problem is bandwidth allocation, which is concerned with successful integration of link capacities through the different types of services. Given that a virtual path is a logical direct link, composed of a number of VCs between any two nodes, the last problem is concerned with how to assign bandwidth to each VP in the network in order to optimize performance for all users. For the aforementioned problems, Saroit Ismail (2000) presents several possible solutions.

DISCARDING POLICIES

In high-speed networks acknowledging individual packets is impractical. Consequently, when congestion builds up and packets have to be dropped, entire messages are lost. For a message to be useful, all packets comprising it must arrive successfully at the destination. Therefore, the problem is which packets to discard so that as many as possible complete messages are delivered, and so that congestion is alleviated or avoided altogether. Lapid et al. (1998) studied selective discarding policies as a means for congestion avoidance and compared these policies to non-discarding policies.

FAST BROADCAST

Traditional broadcast protocols are inappropriate for high-speed networks and are limited by the speed of software processing, which becomes bottlenecked as network speeds increase. Gopal et al. (1999) presented broadcast protocols that are appropriate for high-speed networks and are tolerant of failures involving the loss of messages. The protocols are based primarily on the simple hardware functions present in a high-speed network.

FUTURE TRENDS

Sato (2007) discuss some of the cutting-edge photonic networking technologies including segmented network design, quasi-dynamic network design, multi-layer optical path network design, and optical fast circuit switching. We envisage a rapid growth in the use of IP routers and ATM switches interconnected with SONET links. Backbones will be optical. Consequently, when capacity becomes a bottleneck in the SONET links, wavelength-division multiplexing (WDM) is implemented. Links at 40 Gbps are going to emerge. The future network might expand as follows. ATM switches and IP routers are realized, and all users are connected to ATM switches or IP routers. The ATM switches and IP routers are additionally interconnected with each other by SONET links without any SONET networking. A WDM passive optical network (PON) will be used to add more links as needed and also will provide add-drop multiplexing functions. WDM-PON uses multiple optical wavelengths to increase the upstream and/or downstream bandwidth available to end-users. In general, WDM point-to-point uses different wavelengths to carry information from one switch to another. WDN could potentially drop one (or more) of those wavelengths at the node in order that a certain conductivity is provided between the nodes on this particular wavelength.

Nevertheless, the network is not reconfigurable. In dynamic network configurations, it is achievable to rearrange the add-drop wavelength.

It is worth mentioning that WDM can also be used to provide add-drop multiplexing (ADM) functions in ATM or IP. This future scenario will provide good flexibility and cost effectiveness. Additionally, a high-speed multimedia network will be mainly wireless at the edges with access to a high-speed optical backbone infrastructure, including optical switches. In the extreme, one can envisage a network that consists only of wireless access to an optical backbone, although in practice copper wire will be in use for a very long time. Fourth-generation mobile wireless networks are emerging to offer capacities up to 150 Mbps to fully mobile users in various environments. In such rapidly developing environment, resource management remains an important issue in the future. Efficient resource management is an issue of supreme importance due to: 1) the rapid increase in size of the wireless mobile community; 2) its demand for high-speed multimedia communications; and 3) the limited resources. Iraqi and Boutaba (2007) outline some of the crucial resource management problems in wireless and mobile networks. From another perspective, Kanellopoulos et al. (2006) provide an overview of QoS issues required to support end-to-end QoS guarantees for wireless multimedia communications.

CONCLUSION

This chapter has presented high-speed networking technologies and standards such as ATM, Fast Ethernet, 10 Gigabit Ethernet, SONET, RPR, PBT, PBB, Transport – MPLS, and OTN. It has analysed future trends in high-speed multimedia networking and the crucial problems emerging in this research area. We believe that for the next generation Internet the speed of network links and routers will be improved dramatically so that network congestion will be very unlikely and QoS guarantees will be provided automatically (Gu et al., 2007). This effort includes optical WDM technologies being investigated by Next Generation Internet (NGI) initiative (http://www.ccic.gov/ngi/).

REFERENCES

Allan, D., Bragg, N., McGuire, A., & Reid, A. (2006). Ethernet as carrier transport infrastructure. *IEEE Communications Magazine*, *44*(2), 134–140.

Armitage, G. (2000). MPLS: the magic behind the myths. *IEEE Communications Magazine*, *38*(1), 124–131.

Aurrecoechea, C., Campbell, A., & Hauw, L. (1998). A survey of QoS architectures. *Multimedia Systems*, *6*(3), 138-151.

Chen, W. T., Deng, Y-W., Huang, N. F., Wang, J. S., & Lin, H. C. (1999). On the design of a high-speed network environment for multimedia applications. *Proc. Natl. Sci. Counc. Repub. Of China Part A*, *23*(1), 111-124.

Cheng, J., Yu, H.-C., & Sincoskie, W. (2005). Meeting the broadband access infrastructure demands: The promise of Gigabit Ethernet. *Technological Forecasting & Social Change*, *72*(1), 1-10.

Ciena (2005). *The value of OTN for network convergence and IP/Ethernet Migration* (White paper) Ciena Corporation [Online:] www.xchangemag.com/bgpl/XCciena6.pdf

Davik, F., Yilmaz, M., Gjessing, S., & Uzun, N. (2004). IEEE 802.17 resilient packet ring tutorial. *IEEE Communications Magazine*, *42*(3), 112–118.

Gibson, J. (2000). *Multimedia Communications: Directions and Innovations*. Academic Press, 318 pages.

Giordano, S., Salsano, S., Van den Berghe, S., Ventre, G., & Giannakopoulos, D. (2003). Advanced QoS provisioning in IP networks: The European

Premium IP Projects. *IEEE Communications Magazine*, *41*(1), 30-36.

Gopal, A., Gopal, I., & Kutten, S. (1999). Fast broadcast in high-speed networks. *IEEE/ACM Transactions on Networking*, *7*(2), 262-275.

Gu, Y., Towsley, D., Hollot, C., & Zhang, H. (2007). Congestion control for small buffer high-speed networks. *IEEE INFOCOM 2007* (pp. 1037-1045).

IEEE (1995) Supplement to ISO/IEC 8802-3:1993 (1993). *Local and Metropolitan Area Networks: Media Access Control (MAC) Parameters, Physical Layer, Medium Attachment Units, and Repeaters for 100 Mb/s Operation, Type 100BASE-T (Clauses 21-30).*

IEEE 802.17 (2004). *Resilient packet ring (RPR) access method and physical layer specifications*. IEEE Standard 802.17.

IEEE 802.3 (1998). *Information Technology—Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks—Specific requirements—Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. IEEE, New York.

Iraqi, Y., & Boutaba, R. (2000). Resource management issues in future wireless multimedia networks. *Journal of High Speed Networks*, *9*, 231-260.

ITU-T (2005). Recommendation G.8110. *T-MPLS Layer Network Architecture*. January 2005.

Kanellopoulos, D., Pintelas, P., & Giannoulis, S. (2006). QoS in wireless multimedia networks. *Annals of Mathematics, Computing & TeleInformatics*, *1*(4), 66-75.

Kirstädter, A., Gruber, C., Riedl, J., & Bauschert, T. (2006). Carrier-grade Ethernet for packet core networks. *Proc., SPIE,* 6354. 635414-1–635414-14, Gwangju, Korea, Sept. 2006.

Lapid, Y., Rom, R., & Slidi, M. (1998). Analysis of discarding policies in high-speed networks. *IEEE Journal on Selected Areas in Communications*, *16*(5), 764-777.

Lometti, A. (2006). Transport-MPLS (T-MPLS): An enabling technology for packet transport networks. *Proc. ECOC 2006*, *5,* 91–110.

Lu, G. (2000). Issues and technologies for supporting multimedia communications over the Internet. *Computer Communications*, *23*(14-15), 1323-1335.

Mammeri, Z., & Lorenz, P. (2004). High-speed networks and multimedia communications, *Lecture Notes in Computer Science (LNCS)*, 3079.

Oie, Y., Sueda, T., Murata, M., & Miyahara, H. (1991). Survey of switching techniques in high-speed networks and their performance. *International Journal of Satellite Communications, 9*, 285-303.

Pillalamarri, S., & Ghosh, S. (2005). High-speed networks: definition and fundamental attributes. *Computer Communications*, *28*(8), 956–966.

Saroit Ismail, I. A. (2000). Bandwidth problems in high-speed networks. *IBM J. Res. Develop.*, *44*(6), 919-938.

Sato, K.-I. (2007). Recent developments in and challenges of photonic networking technologies. *IEICE Trans. Commun.*, *E90–B*(3), 454-467.

Sexton, M., & Reid, A. (1997). *Broadband Networking: ATM, SDH, and SONET*. Artech House, Norwood, MA.

Sterbenz, J., & Touch, J. (2001). *High-speed networking: A systematic approach to high-bandwidth low-latency communication.*, New York, US: John Wiley.

KEY TERMS

Bandwidth on Demand: It refers to data rate measured in bit/s (channel capacity or throughput-bandwidth consumption), which is required in order to transfer continuous media data (e.g. video).

Carrier Sense Multiple Access With Collision Detection CSMA/CD: It is a network control protocol in which a carrier-sensing scheme is used. In CSMA/CD, a transmitting data station that detects another signal while transmitting a frame, stops transmitting that frame, transmits a jam signal, and then waits for a random time interval (known as “backoff delay” and determined using the truncated binary exponential backoff algorithm) before trying to send that frame again.

Delay Variation (or Delay Jitter): It is a term used for the variation of end-to-end delay from one packet to the next packet within the same packet stream (connection/flow).

Fibre Channel (FC): It is a gigabit-speed network technology used for storage networking. FC can run on both twisted pair copper wire and fiber-optic cables. Fiber always denotes an optical connection, whereas fibre is always the spelling used in “fibre channel” and denotes a physical connection, which may or may not be optical.

IEEE 802.1ag (Connectivity Fault Management): It specifies support for proactive alarming of service faults and assists with the detection, verification and isolation of connectivity failures.

Media Access Control (MAC): It is a sublayer of the Data Link Layer (DLL) specified in the Open Systems Interconnection Reference Model (OSI-RM). It provides addressing and channel access control mechanisms that make it possible for several terminals or network nodes to communicate within a multipoint network, typically a local area network (LAN) or metropolitan area network (MAN).

Multi Protocol Label Switching (MPLS): It lies between traditional definitions of Layer 2 (Data Link Layer) and Layer 3 (Network Layer). It provides a unified data-carrying service for both circuit-based clients and packet-switching clients, which provide a datagram service model. It can be used to carry many different kinds of traffic, including IP packets, as well as native ATM, SONET, and Ethernet frames.

Operations, Administration, and Maintenance (OAM): This is a general term used to describe the processes, activities, tools, standards, etc involved with operating, administering, and maintaining any networked computer system.

Quality of Service (QoS): QoS functionality enables service providers to guarantee and enforce transmission quality parameters (e.g. bandwidth, jitter, delay, packet loss ratio) according to a specified service-level agreement (SLA) with the customer.

Virtual path (VP): It is defined as a set of multiplexed circuits, which terminate at common end nodes such as switching nodes and network gateways.

Chapter L
Emerging Telecommunications Technologies:
Cognitive Radio

J. Joaquín Escudero-Garzás
University Carlos III de Madrid, Spain

Ana García-Armada
University Carlos III de Madrid, Spain

ABSTRACT

The goal of this chapter is to introduce the novel concept of cognitive radio (CR) for wireless telecommunications. Cognitive radios are a new type of radio devices that include cognition and reconfigurability features. The raising number of studies in different areas of research shows their potential and the expectation created among the telecommunications community. In this chapter, the authors first introduce the reader to the new paradigm that cognitive radio networks have created; more specifically, they explain in detail the new next generation networks. Given that our intention is to introduce cognitive radio, the authors focus on the challenges in PHY layer and MAC sublayer and the most relevant studies in these fields. Finally, the integration of game theory and cognitive radio creates a new paradigm where the advantages of both technologies merge to solve complex problems.

INTRODUCTION

Wireless telecommunications are evolving towards very complex environments due to two main factors: the coexistence of different types of networks, such as WiFi and WiMAX, and the requirements of mobility and multimedia contents demanded by the users. Hence, a major concern is to be aware of and react to the variations in the wireless channel. Until now, the adaptation of wireless devices to the environment has been done through some type of adaptive transmission, but more intelligent functionalities than adaptive modulation are needed. Cognitive radio can deal efficiently with the aforementioned complex problem, endowing radios with the capability of thinking and, in the second stage, acting consequently.

The concept of cognitive radio is relatively new. Cognitive radio is a radio system capable to gather the features of the environment and to adapt their transmissions accordingly. To this end, it is essential to endow a cognitive radio with two primary functions: the capability of capturing information from the environment and reconfigurability.

This chapter is organized as follows: first, a descriptive overview of cognitive radio and some related concepts is presented to the reader. This overview presents relevant and recent works done for PHY and MAC layers for cognitive radio. We conclude this chapter with a discussion on the possible connection between cognitive radio and game theory.

COGNITIVE RADIO

As mentioned in the introduction, the term cognitive radio is credited to Mitola III and Maguire (1999). They introduced cognitive radio as a radio device with cognition functions. Before explaining in detail the concept of cognitive radio, it is useful to introduce the precedent technology, Software Defined Radio (SDR). SDR is defined as (FCC, 2005) "any radio that includes a transmitter in which the operating parameters of frequency range, modulation type or maximum output power can be altered by making a change in software without making any changes to hardware components that affect the radio frequency emissions". In other words, by using software the transmit power, adaptive modulation and transmission frequency, among other transmission parameters and features, may be changed.

Nevertheless, cognitive radio goes a step further than SDR, empowering the radio to change the transmission parameters based on interaction with the environment of operation. Among the wireless scientific community a widely accepted definition of cognitive radio is provided by Haykin (2005); he considers a cognitive radio as "an intelligent wireless communication system that is aware of its surrounding environment", with the ability of adapting its states to environmental changes. In other works, a cognitive radio can be described as a SDR endowed with a "brain" which not simply reacts to external changes but also "thinks" about the actions to take. Two key characteristics must then be associated to any cognitive radio: cognitive capability and reconfigurability.

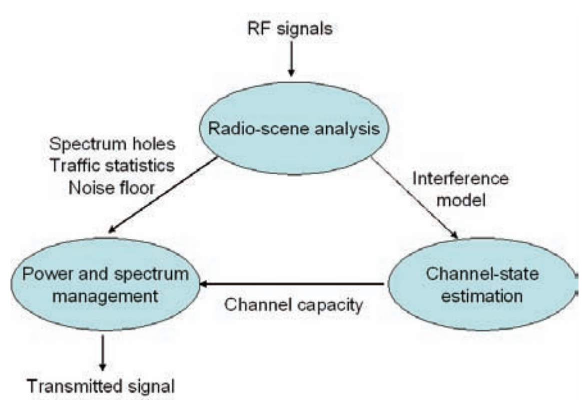
However, what is exactly contained in the cognition part of these radios? Citing again (Haykin, 2005), three major tasks make up the cognitive capability:

1. Radio-scene analysis, in charge of the detection of spectrum holes and the estimation of interference.
2. Channel identification, which carries out the estimation of both channel state information (CSI) and channel capacity.
3. Power and spectrum management.

Figure 1 provides how these three tasks are connected. In practice, cognitive capability may reduce to signal processing or machine learning algorithms for implementation purposes.

Reconfigurability seems to be easier to characterize than cognition capability. This functionality is provided by a SDR platform, which carries out the suitable actions to change the radio parameters

Figure 1. Cognition tasks



values and to adapt the transmitted signal to the new environment. Consequently, the design of SDRs for cognitive radios must consider the dialog between the cognitive part of the radio and the software radio component.

Cognitive radio by virtue of its reconfigurability and cognition properties can be utilized in a wide range of situations: ad hoc, mobile and wireless sensor networks; frequency and time-variant channels; cross-layer design and heterogeneous environments. Since the deployment of cognitive radio-based networks is expected to rise in the near future, the Working Group on Wireless Regional Area Networks ("WRANs") IEEE 802.22 has been created to develop a standard for a Cognitive Radio-based PHY/MAC/air-interface. This standard defines a scenario for non-license devices on a non-interfering basis, by allocating the spectrum already allocated to the TV Broadcast Service to other services (Cordeir et al., 2005).

NEXT GENERATION NETWORKS

NEXT GENERATION NETWORKS (xGN) are a new type of networks very close to the concept of Cognitive Radio. They are also known as *DYNAMIC SPECTRUM ACCESS* (DSA) networks, since nodes will use dynamic spectrum access techniques. The idea of DSA techniques is that wireless nodes have no preallocated bandwidth for their transmission and they shall obtain it dynamically according to the available frequencies. In the case of open spectrum (e.g. ISM band), the spectrum sharing is free and the access rules for a dynamic access are relaxed. However, the most interesting challenge comes when the bandwidth is licensed, situation that, on the other hand, is much more frequent than the open spectrum.

Historically spectrum has been considered as a scarce resource for wireless communications and, therefore, the governments have traditionally controlled the spectrum management and assigned it on a service basis. Nevertheless, rather than a problem of scarcity of suitable frequencies, it is a problem of efficiency: in licensed wireless networks, the percentage of time that licensed users (or licensees) effectively utilize the spectrum is quite low (Chiang, Rowe and Sowerby, 2007), leading to a situation in which the efficient utilization of the licensed frequencies is actually poor.

Two types of users coexist in NEXT GENERATION NETWORKSs: *primary* or *licensed users* have a license to operate in a certain frequency band and will always transmit in their licensed band. On the contrary, *secondary users* have no license to operate and make dynamic use of the spectrum, provided that suitable frequency bands are available. Logically, secondary users have an additional commitment: they must not interfere with licensed users. We note that for DSA, secondary users may scan both licensed and non-licensed spectrum bands and may transmit in both types of bands. Thus, we can clearly distinguish two different cases, licensed and unlicensed spectrum. We will devote the following subsections to analyze in detail both cases.

Licensed spectrum. For licensed spectrum, the context of xGN provides two types of scenarios depending on the policy adopted with respect to DSA: spectrum underlay and spectrum overlay.

Spectrum underlay policies allow the secondary users to operate sharing with primary users the same frequency bands; therefore, the interference secondary users provokes on primary users must be under a certain threshold.

Spectrum overlay policies are used to carry out opportunistic spectrum access, transmitting only using certain frequency bands called *spectrum holes* or *white spaces*. Spectrum holes are temporally not occupied by primary users, and therefore available for secondary users. For spectrum overlay, or opportunistic spectrum access in general, cognitive radios act as follows: firstly, *SPECTRUM SENSING* allows the detection of the frequency bands that are not being used in a particular instant of time and place, so determining the spectrum holes. In a second step, *spectrum analysis* provides the quality of each spectrum hole. Lastly, *spectrum decision* is done to choose the most suitable spectrum hole to carry out the transmission.

The implications in the spectrum management are significantly different depending on whether spectrum underlay or spectrum overlay is employed, hence the strategies to tackle the problem of DSA cannot be the same. For instance, for spectrum underlay PHY layer design must be done very carefully, because the transmissions of secondary user can interfere with primary users only within certain tolerable limits. An appropriate modulation in this case would be Ultra Wide Band (UWB), and we will discuss it and other modulation schemes in a subsequent section. In other cases, a secondary user may have some knowledge about primary user's information; therefore, he might also benefit in information-theoretic terms: a coding technique as dirty paper coding may be used to mitigate interference.

Unlicensed spectrum. When considering unlicensed bands for spectrum access, the discussion is quite different because of the fairness among all the users and the non-existence of "VIP" users. Therefore, all users are equal with respect to their rights to access the spectrum. This leads to a completely different treatment of spectrum sensing for unlicensed and licensed bands: secondary users having the possibility of accessing both types of bands must implement both types of sensing. Furthermore, the utilization of both licensed and unlicensed bands implies the design of medium access protocols solving jointly spectrum access for the two cases.

PHY AND MAC COGNITIVE RADIO

In the following, we will focus our attention on PHY and MAC aspects for cognitive radios, and for each of them we will provide a detailed summary of the state-of-the-art research. Power management and adaptive modulation are the most relevant subjects in PHY layer and receive more attention than other topics. Related to MAC sublayer, we believe that the most important issues are the detection of users and the utilization of the spectrum; this justifies that the majority of our discussion on MAC sublayer is devoted to them.

PHY LAYER

Cognitive radio has been identified as a very adequate solution for PHY layer dynamic problems such as power management or adaptive modulation and coding, given that the configuration must be real-time adapted according to the surrounding conditions. The promising results obtained, for instance, by using power management or UWB for cognitive radios allow us to envisage a vast use of "cognitivity" in the PHY layer design. Although these immediate applications of cognitive radio are important, cognitive radio networks are the most challenging framework, since all devices perform the cognitive function and they do not act separately.

Power Management

In most wireless networks a central node, e.g. a base station or access point, controls the transmit power of the devices and its own power in order to achieve the desired performance. In other cases such as wireless sensor networks, the power management can be performed in a distributed manner for a faster arrangement of the network. Nevertheless, for cognitive radio networks power management must cope with both situations, given that a cognitive radio may completely ignore the nature and topology of the networks that share the same geographical space. Moreover, the consideration of power management along with other issues, such as interference and capacity issues, can provide many advantages.

In cognitive networks, power allocation presents an additional constraint: in licensed spectrum, the coexistence of primary and secondary users leads to a limitation in the transmission power of secondary users, since they should not cause any interference degrading the primary users' performance. This implies that power allocation must be done carefully looking at the interference temperature constraints imposed by the primary users, as can be seen in (Bansal, Hossain and Bahrgava, 2007; Wang, Peng and Wang, 2007). However, power allocation performing jointly with spectrum sensing may achieve an improvement in the algorithms.

Spectrum sensing is not limited to the activity of primary users: when applied to secondary user's transmissions, spectrum sensing provides information to the system about interference among secondary users. Consequently, not only the interference constraints imposed with regard to primary users are taken into account. Joint admission control and power allocation also promises to be an effective manner to address power management for cognitive networks, as shown by Zhang, Liang and Xin (2007).

Adaptive Modulation

Cognitive radios are aware of what is happening in their environment by making use of their cognition capability; at the same time, they use their reconfigurability functionality to adapt their transmission to their current conditions. To carry out adaptive transmission, adaptive modulation appears to be the immediate option (Gao, Qian, Vaman and Qu, 2007). Nevertheless, other modulation techniques are capable to customize the values of the transmission parameters and have demonstrated that they can perform satisfactorily in cognitive radio networks. Multicarrier modulation OFDM (Orthogonal Frequency Division Multiplexing) appears as one of the most popular modulations to be employed in cognitive networks (see as examples (Rajbanshi, Wyglinski and Minden, 2006; Budiarjo, Nikookar and Lighthart, 2006)), as well as the derived medium access technique OFDMA (Orthogonal Frequency Division Multiple Access). The basic idea of OFDM is to divide the spectrum into different subbands, which are kept orthogonal thanks to a cyclic prefix in non-ideal propagation conditions. When using OFDM the system adaptively allocates subcarrier bands, data rate and power to each user. It has also demonstrated its efficiency against inter symbol interference (ISI) and multipath effects.

The mentioned advantages have entailed OFDM to be adopted as the modulation of different wireless standards (i.e. IEEE 802.11a, IEEE 802.16) and to be considered as a candidate for PHY layer of the IEEE 802.22 cognitive networks standard. Nevertheless, the majority of OFDM schemes assume that resource allocation is known a priori with the aid of a central node or base station. This assumption is not realistic in DSA scenarios due to the very changing nature of the bandwidth allocation processes, but joint detection and decoding schemes may result very effective to overcome this problem (Li, Mow, Lau, Siu, Cheng, and Murch, 2007).

Spread Spectrum

Spread spectrum (SS) techniques are well positioned for wireless communications in general, given that they mitigate the effect of intersymbol and narrowband interference. We will specifically address a type of SS modulation called Code Division Multiple Access (CDMA). SS is based on the transmission of signals with a much larger bandwidth than necessary, with the aim of hiding the signal below the noise floor. When using CDMA, each user has his own spreading code to multiply the data signal. Up to now, CDMA-based cognitive networks are in the first steps, despite of their potential benefits (Qu, Milstein and Vaman, 2008).

Ultra Wide Band (UWB)

UWB is defined as any radio communications technique using very broadband signals: the utilized bandwidth exceeds at least 20% the value of the frequency carrier, with a minimum bandwidth of 500 MHz. Among the most valuable features of UWB we may find the very low power consumption and the capacity to allocate power adaptively; these are the reasons why UWB is very well positioned for its use in underlay cognitive networks, where the main goal is to minimize the interference upon primary users. The resistance of UWB techniques in presence of multipath and simple transceiver circuitry also contribute to advocate for the future development of UWB-based cognitive networks and the study of different applications with UWB techniques (see (Sahin, Ahmed and Arslan, 2007) and references therein).

Cooperative Diversity for Cognitive Radio Networks

Cooperative transmission, whether in its basic form of relaying or its more complex schemes, has demonstrated to be very effective in wireless networks (Sendonaris, Erkip and Aazhang, 2003). The simplest scheme is *Amplify-and-Forward* (AF), and the intermediate node acts as a relay, amplifying the received signal according to his own power constraints. In *Decode-and-Forward* (DF) schemes, the intermediate node performs detection and decoding for the received signal and encodes the signal again, reducing the impact of the noise. While AF and DF schemes are fixed, *Selection-and-dynamic* schemes allow the cooperating terminals to adapt their transmission according to the channel coefficients.

Simeone, Gambini, Bar-Ness and Spagnolini (2007) have realized the usefulness of cooperative diversity for cognitive radio networks, coining the term *cognitive relay* as a means of helping primary users to relay their information. Interestingly, they show that secondary users receive also some benefit.

SPECTRUM SENSING

The general detection problem consists in determining whether the joint presence of a signal *A* plus noise *w*[n] or only noise *w*[n] is present. Mathematically, for the simplest case the signal to be detected is $A = 1$. Thus, the detection problem reduces to decide between two hypothesises H_0 and H_1: H_0 stands for only noise is present and H_1 that signal plus noise is present. The noise *w*[n] is distributed according to a Gaussian probability density function (PDF) of mean $\mu = 0$ and variance σ^2 (in short notation $w[n] \sim N(0, \sigma^2)$). With the precedent characterization, the formulation of the detection problem is:

$$H_0 : x[0] = w[0] \text{ if } x[0] < \tfrac{1}{2}$$
$$H_1 : x[0] = 1 + w[0] \text{ if } x[0] > \tfrac{1}{2}$$

Associated to the detection problem, two probabilities characterize the performance of the detector: the *probability of false alarm* p_{FA} gives the probability of deciding a signal is present when it is not, and the *probability of miss detection* p_{MD} gives the probability of deciding a signal is not present when in fact it is. It is straightforward to realize that p_{FA} represents the probability of deciding H_1 when H_0 is true and p_{MD} represents the probability of deciding H_0 when H_1 is true. Formally:

$$p_{FA} = \Pr\{H_1; H_0\} = \Pr\{ x[0] > \tfrac{1}{2}; H_0\}$$
$$p_{MD} = \Pr\{H_0; H_1\} = \Pr\{ x[0] < \tfrac{1}{2}; H_1\}$$

In cognitive radio networks, compared to the other wireless networks, we find very different circumstances for spectrum sensing. Since the wish is to have very reliable detectors, the objective in the detection problem shall be to minimize both probabilities. This desire means for spectrum sensing a twofold problem: the detection, mainly of primary users, must be done by both ends of the communication, transmitter and receiver. Additionally, fast and reliable detection of primary users is a must; however, we find a very important difficulty in the way: typical environments have channels with very low signal-to-noise ratio (SNR), yielding very unfavourable conditions for detection. Moreover, if not enough, cognitive radios must deal with the hidden and exposed terminal problems. The *hidden node problem* refers to terminals that can communicate to a base station but cannot hear each other, whether due to the presence of obstacles or for being out of their respective transmission ranges. This situation originates that in carrier sense multiple access (CSMA) nodes assume they can transmit but in fact cannot, causing a collision (Figure 2). The *exposed terminal problem* takes place when a transmitter is prevented from sending packets due to the existence of another transmitter in the nearby: if BS1 wants to transmit to U1, it must detect whether BS2 is transmitting to U2, since they are exposed to each other transmissions (Figure 3). The traditional receiver approaches, namely matched filter and energy detectors, do not suit well to the described paradigm, as we explain in the following.

Matched filter receivers require a previous knowledge about the signal to be detected, e.g. modulation type and signal shape. In coherent

Figure 2. The hidden terminal problem

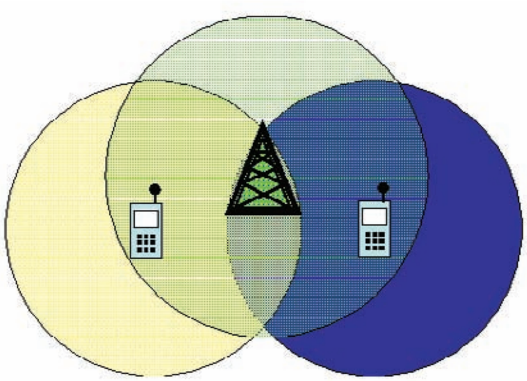
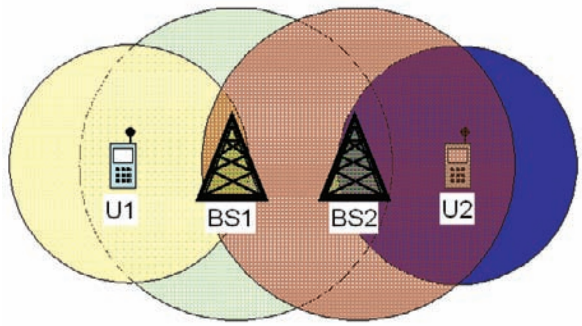


Figure 3. The exposed terminal problem



detection systems information contained in pilots or preambles may help the receiver, yet in this case the privacy of users can be seriously compromised. Moreover, in cognitive radio networks a receiver should have as many matched filters as possible bands of utilization if we use different standards in these bands.

Energy detectors monitor the received energy. Despite of their ease of implementation, they are restricted to systems where certainty about the noise power and the strength of the signal of interest are sufficiently high to allow a correct performance of the detector. Moreover, in cognitive radio scenarios SNR is frequently very low, provoking the impossibility of determining the presence or absence of another user, either primary or secondary.

Once we have shown why energy and matched filter detection are not suitable for cognitive radios, we explore the possibility of ciclostationarity receivers. They make use of the ciclostationarity characteristic, which is present in some signals: a signal is *ciclostationary* if we can find a non-linear transformation of the signal that generates spectral lines. The main advantage of these receivers is that the autocorrelation function of the signal, which results to be periodic, completely characterizes the signal. Furthermore, by using ciclostationarity the detection of signals without any previous knowledge about them can be accomplished, as well as the extraction of features such as bit-rate, as shown in the pioneer Gardner's work (1986).

Very interesting works have continued the same line (see for example (Gelli, Izzo and Paura, 1996; Chevreuil, Loubation and Vanderdorpe, 2000)), yielding the basis for the design of cicloestationarity-based receivers for cognitive radio networks, such as (De Angelis, Izzo, Napolitano, and Tanda, 2005; Li, Bar-Ness, Abdi, Somekh and Su 2006; Tkachenko, Cabric, and Brodersen, 2007). Nevertheless, practical implementation is a key issue for this type of receivers, since it requires long periods of observation and is computationally intensive.

Therefore, cognitive radios need new approaches facing the difficulties mentioned thus far, as done by Wild and Ramchandran (2005) and De (2007). One of the main drawbacks to carry out reliable spectrum detection appears if a single detector has to perform in very unfavourable situations, such as the detection of primary users with low SNR: the detection will probably become unfeasible. This unfeasibility led to Cabric et al. (2004) to envisage cooperation as the future for sensing techniques in cognitive networks, having their work a continuation in (Cabric, Tkachenko and Brodersen, 2006), where the authors propose a cooperative detection scheme with higher performance than individual detectors.

The abovementioned cooperative sensing technique must consider how to decide about the sensing result. The first option is to centralize the decision: the fusion centre evaluates the individual decisions through a fusion function and takes the final sensing decision (see for instance (Visotsky, Kuffner and Peterson, 2005) and (Sun, Zhang amd Leitaef, 2007)). Nevertheless, not always centralized approaches are possible: distributed sensing techniques, making use of the exchange of information among the nodes, allow to decide on sensing without the participation of any fusion centre.

Cooperative sensing may also be accomplished distributedly. Distributed sensing has been intensively applied to wireless sensor networks (Chen, Tong and Varshney, 2006; Lin, Chen and Suter, 2007), but the well-known limitations in energy and processing capabilities condition absolutely the solutions. Therefore, although envisaged to be very helpful for cognitive radio (Uchiyama et al., 2007; Gandetto and Regazzoni, 2007), there is a long road ahead for the study of distributed sensing for cognitive radios.

A third way to carry out cooperative sensing appears to emerge with the application of cooperative diversity to spectrum sensing (Mishra et al, 2006; Ganesan and Li, 2007a and 2007b), since the cooperation may reduce the detection time by making use of AF protocol and may also increase the overall agility of the network.

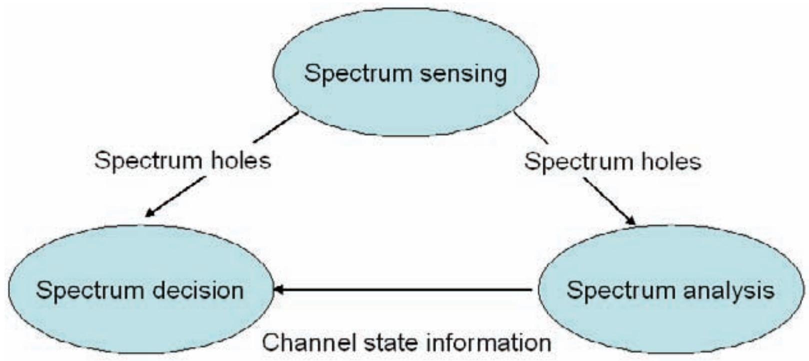
Given that receivers perform sensing operation repeatedly, it is of paramount importance that spectrum sensing is as energy-efficient as possible, mainly in battery-powered devices. Therefore, one of the future challenges for cognitive networks is the study of spectrum sensing solutions (for instance, design of suitable receivers or transmission schemes) capable of minimizing energy consumption.

SPECTRUM ACCESS

One of the main challenges in cognitive radio networks is the *SPECTRUM ACCESS*. On the one hand, there is a wide range of available spectrum for transmission; on the other hand, two types of users with different privileges coexist and share the spectrum. These are the two main reasons to consider new spectrum access techniques in these networks. Moreover, they must take into account the use of DSA. We note that, related to spectrum issues for cognitive radios, in the literature we can find the term *spectrum agility* to characterize that devices can opportunistically use a range of frequencies among various frequency bands.

In order to describe SPECTRUM ACCESS for cognitive radio networks, we will follow the process shown in Figure 4. In the first stage *SPECTRUM SENSING* (seen in the precedent sec-

Figure 4. Spectrum access in cognitive radios



tion) detects the spectrum holes; next *spectrum analysis* provides the characterization of the idle bands and, finally, the *spectrum decision* is taken considering information from both spectrum sensing and spectrum analysis. We detail each of these stages in the following paragraphs.

Spectrum analysis has the scope of estimating the characteristics of the detected spectrum holes, whose suitability for a certain transmission is evaluated. This evaluation consists in determining the quality of the spectrum holes by means of parameters such as interference level, channel error rate and path loss; therefore, the knowledge of information about the channel, known as *channel-state information* (CSI), is required in a centralized manner. In brief, *channel estimation* is essential to perform spectrum decision and, consequently, for SPECTRUM ACCESS.

When dealing with channel estimation, pilots transmission has performed quite satisfactorily in traditional wireless networks. Nevertheless, efficient channel estimation in cognitive radio networks implies different challenges such as a continuous channel tracking for the detection of new primary users and estimation in channels with very low SNR. Since very efficient use of power and bandwidth is also required, the use of pilots may not be the best solution and alternatives such as particle filtering have an opportunity (Haykin et al., 2004). A related work to this topic is (Rondeau et al., 2004), where the authors describe a channel model for cognitive radios.

As part of the CSI, the estimation of channel capacity is a very important aspect for the final decision about which of the spectrum holes should be used: if the chosen spectrum hole does not provide sufficient capacity for the transmission, it becomes useless despite of the adequate values of the remainder of CSI parameters. Thus far, the capacity problem is oriented to underlay networks: information from primary transmitters is applied to perform coding (dirty paper and Gel'fand-Pinkser coding are the preferred ones) with the aim of reducing interference and the channel is modelled mostly as a Gaussian interference channel (Devroy et al., 2005; Jovicic and Viswanath, 2006; Devroye et al., 2006). Recently, a first attempt in modelling the throughput for the overlay case (called here "interwave") has appeared based on the new concept of "two switch" channel model (Jafar and Srinivasa, 2007; Srinivasa and Jafar, 2007b).

The final stage of the SPECTRUM ACCESS process is the *spectrum decision*. Among all the spectrum holes, the most appropriate spectrum band is selected: the quality requirements are compared to the estimated quality of the spectrum holes, and so providing the most suitable band.

A high percentage of the works in cognitive networks is devoted to SPECTRUM ACCESS, providing very interesting approaches in the form of MAC protocols that assemble proposed solutions for each of the aforementioned spectrum access stages. The proposals of Vuran and Akyildiz (2007) and Chou, Sai, Hyoil and Shin (2007) may serve as examples, providing the latter a more complete framework describing cross-layer protocols involving PHY and MAC layers.

We can consider *spectrum sharing* techniques under the spectrum access umbrella, facing a more specific situation: these techniques allow a fair sharing of the spectrum holes among the cognitive radios. Again, the fact of operating with DSA, either licensed or unlicensed spectrum, leads to a completely new situation. For instance, among other issues, the amount of interference must be considered both for licensed and unlicensed users (Srnivasa and Jafar, 2007b; Atakan and Akan, 2007); or the problem arose if several service providers share the bandwidth within the same area (Dimitrakopoulos and Demestichas, 2005; Acharya and Yates, 2007).

The previous sections stated, beyond all doubt, the importance of very accurate detectors when secondary users need to identify spectrum holes. Furthermore, they must detect any primary user accessing the channel and act accordingly, even vacating the channel if needed and performing a sort of "handoff" in order to hold the communication. This topic is one of the most challenging when dynamic spectrum access is being used, and by Fu, Zhou, Xu and Song (2007) discuss the major concerns about handoff in Cognitive Radio.

INTERFERENCE

Interference is always present in wireless networks; even more, some telecommunication networks are limited by interference. Cognitive radios can target on minimizing interference, for example, by choosing the adequate modulation, adapting transmit power or by means of bandwidth allocation.

The problem of interference can be addressed for cognitive radios in the framework of two different approaches: the interference cancellation, or minimization model, and the INTERFERENCE TEMPERATURE model. For the former, the focus is put on transmissions and they are controlled to avoid interfering emissions; the existing literature on this topic is vast and the proposals are many (see (Popovski, Yomo, Nishimori, Di Taranto and Prasad, 2007) as an example). For the latter, the problem is complicated to solve, since it implies to measure the interference temperature, and to carry out effectively this measurement is difficult. Hence, although in the recent literature there are approaches in modelling cognitive radio networks with interference temperature constraints as (Wang, Cui, Peng and Wang, 2007) and (Bater, Tan, Brown abd Doyle, 2007), reliable methods to determine the interference temperature need to be developed.

Now, we will focus on the second approach. For the interference temperature model some but limited interference level is allowed in the system: from the point of view of the targeted receiver, as long as interfering signals (i.e. users) appear, they contribute to the noise in the system and the noise floor increases. The concept of *INTERFERENCE TEMPERATURE* is clearly the key and it is defined as (FCC, 2004) "a measure of the RF power generated by undesired emitters plus noise sources that are present in a receiver system (I+N) per unit of bandwidth. More specifically, it is the temperature equivalent of this power measured in units of "Kelvin" (K)." From this definition, the INTERFERENCE TEMPERATURE T_{int} can be calculated as $T_{int}(K) = \frac{I+N}{KB}$. This concept allows to classify an emission as interference according to an straightforward criterion: for each band there exists a pre-established limit value of T_{int}, namely T_{max}; then a transmission is considered to cause interference if the noise floor is larger than the noise associated to T_{max}. Being N_{floor} the noise floor when no interfering signal is present in the receiver, we can express the criterion mathematically as follows:

$$N_{floor} + TK_{\text{int}} > TK_{\max}.$$

The challenge in this model is precisely the characterization of the interference temperature, frequently being a complex task involving computationally intensive algorithms.

COGNITIVE RADIO MEETS GAME THEORY

Game theory is a mathematical tool applied successfully for wireless networks if an interaction among the participants (e.g. users, nodes, service providers or access points) exists. In the context of game theory, the participants, named *players*, must decide how to act to satisfy their benefit but also in the interest of the overall network gain. This is exactly what happens in cognitive radio: secondary users attempt to access the network and to do so, they must adapt their transmissions subject to constraints in the form of, for instance, potential interference over primary and secondary users, or achievable throughput.

Up to now, the interest of cognitive radio in game theory has emerged in three areas: power management, spectrum sharing and spectrum access. Nevertheless, our opinion is that a significant part of the work made using game theory for wireless communications may have continuation in cognitive radio networks.

The aim of power games is to allocate power to the users to maximize their benefit, for instance in the form of bit-rate, considering they would transmit with the maximal power if they would act selfishly. The works described in (Wang et al, 2007; Shilun and Zhen 2007) show two interesting approaches to the problem of power allocation to secondary users in licensed spectrum, while Clemens and Rose (2005) address the same problem in unlicensed spectrum.

The characteristic of spectrum games is clear: the scarcity of the resource "bandwidth" establishes a conflict of selfish interests among all users. Next, we provide two examples of Game Theory applied to spectrum access. In the first one, Zhang and Tian (2007) propose a DSA game with complete information where the cognitive radios decide the transmission power spectral density based on CSI. This CSI is available to users by means of a common control channel. The game has the final goal of lightening the communication load in the control channel by using extended Kalman filtering, leading to algorithms for channel tracking that save computational load and reduce the communication overhead in the control channel. In the second example, Nie and Comaniciu (2006) analyze the behaviour of cognitive radios for distributed adaptive channel allocation. They formulate a potential game, a type of games where the utility function of the players behaves like a potential function.

Spectrum sharing games have been addressed for efficient, flexible and fair use of the spectrum, although their extension to cognitive networks is not trivial. However, the approaches of Etkin, Parekh and Tse (2007) and Niyato and Hossain (2007) show the potential of spectrum sharing games. It is envisioned that for next generation networks the cognitive radios could provide more efficient schemes solving open problems such as energy-efficient spectrum sharing.

FUTURE TRENDS

In the context of cognitive radio, some relevant works have been carried out as we have described in the preceding sections, mostly focused on spectrum access and a few topics related to PHY layer (power management, spectrum sensing and modulation schemes such as OFDM). Nevertheless, distributed schemes for spectrum sensing and energy-efficient power control can serve as examples for which many findings are to be discovered. Apart from the mentioned issues above, some remaining problems are open, and an important part of them may fit game theory framework. We would like to emphasize one of the most challenging issues in cognitive radio: the FCC has considered this topic as one of the most important for efficient use of the spectrum in the future; however, there is a notorious lack of accurate measurement methods for reliably obtaining the interference temperature.

CONCLUSION

The management of wireless communications is becoming more and more challenging. The increasing complexity of services and the continuous variations of the environment are creating many new situations which are difficult to address with conventional radio techniques. We have presented cognitive radio as a new concept of radio devices that can respond to this new paradigm. The interaction with this intricate and dynamic environment is the crucial point of these radios: the information obtained is used to adapt the transmissions to the surrounding conditions by the joint action of the reconfigurability and the cognition functions. We have shown that users may be classified as primary and secondary, and the allocation of the licensed bandwidth may not be privative to the licensed users and can be shared among all users under certain conditions. This type of bandwidth management changes completely the design of PHY layer and medium access, and cognitive radio-based approaches provide very interesting solutions. For instance, the design of PHY layer using cognitive radio allows us to solve problems related to power management and adaptive modulation, among others. The importance of reliable spectrum sensing is clear when the detection involves different types of users. Consequently, a wide range of new situations emerge such as the consideration of detection in both transmitter and receiver. In the general context of cognitive radio, ciclostationarity may provide adequate receivers; however, cooperative sensing seems to be preferred since channels with very low SNR are quite common.

A meaningful issue in future networks is spectrum access. When using cognitive radios, we need to determine the available spectrum holes and to estimate the channel, before taking the final decision about which band should be used. Hence, although the spectrum access process specified may become complex, it has shown its efficiency in the design of adequate MAC protocols.

Finally, a detailed study on the features of cognitive radio networks may open the door to the joint framework of cognitive radio and game theory, the framework in which the benefits of the participants are harmonized through the cognitive function of the radios.

REFERENCES

Acharya, J., & Yates, R. D. (2007). A framework for dynamic spectrum sharing between cognitive radios. *IEEE International Conference on Communications,1*, 5166-5171.

Atakan, B., & Akan, O. B. (2007). BIOlogically-Inspired Spectrum Sharing in Cognitive Radio Networks. *IEEE Wireless Communications and Networking Conference, 1*, 43-48.

Bansal, G., Hossain, J., & Bhargava, V. K. (2007). Adaptive power loading for OFDM-based cognitive radio systems. *IEEE International Conference on Communications, 1*, 5137-5142.

Bater, J., Tan, H.-P., Brown, K. N., & Doyle, L. (2007). Modelling interference temperature constraints for spectrum access in cognitive radio networks. *IEEE International Conference on Communications,1*, 6493-6498.

Budiarjo, I., Nikookar, H., & Ligthart, L.P. (2006). Combined spectrum pooling and adaptive bit loading for cognitive radio OFDM-based system.

Symposium on Communications and Vehicular Technology, 1, 73-76.

Cabric, D., Mishra, S. M., & Brodersen, R. W. (2004). Implementation issues in spectrum sensing for cognitive radios. *38th Asilomar Conference on Signals, Systems and Computers, 1*, 772-776.

Cabric, D., Tkachenko, A., & Brodersen, R. W. (2006). Spectrum sensing measurements of pilot, energy, and collaborative detection. *Military Communications Conference, 1*, 1-7.

Chen, B., Tong, L., & Varshney, P. K. (2006). Channel-aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine, 23*(4), 16-26.

Chevreuil, A., Loubation, P., & Vandendorpe, L. (2000). Performance of general transmitter induced cyclostationarity precoders: analysis based on a MMSE-DF receiver. *IEEE Transactions on Signal Processing, 48*(11), 3072-3086.

Chiang, R. I. C., Rowe, G. B., & Sowerby, K. W. (2007). A quantitative analysis of spectral occupancy measurements for cognitive radio. *IEEE 65th Vehicular Technology Conference, 1*, 3016-3020).

Chou, C.-T., Sai, S. N., Hyoil, K, & Shin, K. G. (2007). What and how much to gain by spectrum agility*? IEEE Journal on Selected Areas in Communications, 25*(3), 576-588.

Clemens, N., & Rose, C. (2005). Intelligent power allocation strategies in an unlicensed spectrum. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 37-42.

Cordeiro, C., Challapali, K., Birru, D., & Shankar, N. S. (2005). IEEE 802.22: the first worldwide wireless standard based on cognitive radios. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 328-337.

De, P. (2007). New methods for sensing bandlimited signals in cognitive radio. *IEEE 65th Vehicular Technology Conference, 1*,1916-1920.

De Angelis, V., Izzo, L., Napolitano, A., & Tanda, M. (2005). Cyclostationarity-based parameter estimation of wide-band signals in mobile communications. *IEEE/SP 13th Workshop on Statistical Signal Processing, 1*,1038-1041.

Devroye, N., Mitran, P., & Tarokh, V. (2005). Cognitive multiple access networks. *Proceedings of IEEE International Symposium on Information Theory, 1*, 57-61.

Devroye, N., Mitran, P., & Tarokh, V. (2006). Achievable rates in cognitive radio channels. *IEEE Transactions on Information Theory, 52* (5), 1813-1827.

Dimitrakopoulos, G., & Demestichas, P. (2005). Spectrum exchange in a reconfigurable radio context. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 676-679.

Etkin, R., Parekh, A., & Tse, D. (2007). Spectrum sharing for unlicensed bands. *IEEE Journal on Selected Areas in Communications, 25*(3), 517-528.

FCC Federal Communications Commission (2004). ET Docket No. 03-237. *Establishment of an interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed, mobile and satellite frequency bands.*

FCC Federal Communications Commission (2005). ET Docket No. 03-208. *Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies.*

Fu, X., Zhou, W., Xu, J., & Song, J. (2007). Extended mobility management challenges over cellular networks combined with cognitive radio by using multi-hop network. *8th ACIS International*

Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 1, 683-688.

Gandetto, M., & Regazzoni, C. (2007). Spectrum sensing: A distributed approach for cognitive terminals. *IEEE Journal on Selected Areas in Communications, 25*(3), 546-557.

Ganesan, G., & Li, Y. (2007). Cooperative spectrum sensing in cognitive radio Part I: Two user networks. *IEEE Transactions on Wireless Communications, 6*(6), 2204-2213.

Ganesan, G., & Li, Y. (2007). Cooperative spectrum sensing in cognitive radio Part II: Multiuser networks. *IEEE Transactions on Wireless Communications, 6*(6), 2214-2222.

Gao, S., Qian, L., Vaman, D. R., & Qu, Q. (2007). Energy efficient adaptive modulation in wireless cognitive radio sensor networks. *IEEE International Conference on Communications, 1*, 3980-3986.

Gardner, W. A. (1986). The spectral correlation theory of cyclostationary time-series. *Signal Processing, 11*, 13-36.

Gelli, G., Izzo, L., & Paura, L. (1996). Cyclostationarity-based signal detection and source location in non-Gaussian noise. *IEEE Transactions on Communications, 44*(3), 368-376.

Haykin, S., Huber, K., & Chen, Z. (2004). Bayesian sequential state estimation for MIMO wireless communications. *Proceedings of IEEE (Special Issue on Sequential State Estimation), 92*(3), 439-454.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications, 23*, 201-220.

Jafar, S. A., & Srinivasa, S. (2007). Capacity limits of cognitive radio with distributed and dynamic spectral activity. *IEEE Journal on Selected Areas in Communications, 25*(3), 529 – 537.

Jovicic, A., & Viswanath, P. (2006). Cognitive Radio: An information-theoretic perspective. *IEEE International Symposium on Information Theory, 1*, 2413-2417.

Li, H., Bar-Ness, Y., Abdi, A., Somekh, O. S., & Su, W. (2006). OFDM modulation classification and parameters extraction. *1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 1*, 1-6.

Li, T., Mow, W. H., Lau, V. K. N., Siu, M., Cheng, R. S., & Murch, R. D. (2007). Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference. *IEEE Journal on Selected Areas in Communications, 25*(3), 566-575.

Lin, Y., Chen, B., & Suter, B. (2007). Robust binary quantizers for distributed detection. *IEEE Transactions on Wireless Communications, 6*(6), 2172-2181.

Mishra, S. M., Sahai, A., & Brodersen, R. W. (2006). Cooperative sensing among cognitive radios. *IEEE International Conference on Communications, 1*, 1658-1663.

Mitola, J., III, & Maguire, G. Q., Jr. (1999). Cognitive radio: Making software radios more personal. *IEEE Personal Commununications, 6*(4), 13-18.

Nie, N., & Comaniciu, C. (2006). Adaptive channel allocation spectrum etiquette for cognitive radio networks. *Springer Mobile Networks and Applications, 11*(6), 779-797.

Niyato, D., & Hossain, E. (2007). A game-theoretic approach to competitive Spectrum sharing in cognitive radio networks. *IEEE Wireless Communications and Networking Conference, 1*, 16-20.

Popovski, P., Yomo, H., Nishimori, K., Di Taranto, R., & Prasad, R. (2007). Opportunistic interference cancellation in cognitive radio systems. *2nd IEEE International Symposium on New Fron-*

tiers in Dynamic Spectrum Access Networks, 1, 472-475.

Qu, Q., Milstein, L. B., & Vaman, D. R. (2008). Cognitive radio based multi-user resource allocation in mobile ad hoc networks ssing multi-carrier CDMA modulation. *IEEE Journal on Selected Areas in Communications*, *26*(1), 70-82.

Rajbanshi, R., Wyglinski, A. M., & Minden, G. J. (2006). Adaptive-mode peak-to-average power ratio reduction algorithm for OFDM-based cognitive radio. *IEEE 64th Vehicular Technology Conference, 1,* 1-5.

Rondeau, T. W., Rieser, C. J., Gallagher, T. M., & Bostian, C. W. (2004). Online modelling of wireless channels with Hidden Markov Models and channel impulse responses for cognitive radios. *IEEE MTT-S Digest*, 739-742.

Sahin, M. E., Ahmed, S., & Arslan, H. (2007). The roles of Ultra Wideband in cognitive networks. *IEEE International Conference on Ultra-Wideband, 1,* 247-252.

Sendonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity. Part I. System description. *IEEE Transactions on Communications, 5*(11), 1927-1938.

Shilun, C., & Zhen, Y. (2007). Energy-efficient power control game for cognitive radio systems. *8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 1,* 526-530.

Simeone, O., Gambini, J., Bar-Ness, Y., & Spagnolini, U. (2007). Cooperation and Cognitive Radio. *IEEE International Conference on Communications,.1,* 6511-6515.

Srinivasa, S., & Jafar, S. A. (2007a). The throughput potential of cognitive radio: A theoretical perspective. *IEEE Communications Magazine, 45*(5), 73-79.

Srinivasa, S., & Jafar, S. A. (2007b). Cognitive radio networks: How much spectrum sharing is optimal? *IEEE Global Communications Conference, 1*, 3149-3153.

Sun, C., Zhang, W., & Letaief, K. B. (2007). Cluster-based cooperative spectrum sensing in cognitive radio systems. *IEEE International Conference on Communications*, *1*, 2511-2515.

Tkachenko, A., Cabric, D., & Brodersen, R. W. (2007). Cyclostationary feature detector experiments using reconfigurable BEE2. *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, .1,* 216-219.

Uchiyama, H., Umebayashi, K, Kamiya, Y., Suzuki, Y., & Fujii, T. (2007). Study of cooperative sensing in cognitive radio based ad-hoc network. *18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1,* 1-5.

Visotsky, E., Kuffner, S., & Peterson, R. (2005). On collaborative detection of TV transmissions support of dynamic spectrum sharing. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, *1*, 338-345.

Vuran, M. C., & Akyildiz, I. F. (2007). A-MAC: Adaptive Medium Access Control for next generation wireless terminals. IEEE/ACM Transactions on Networking, *15*(3), 574-587.

Wang, W., Cui, Y., Peng, T., & Wang, W. (2007). Noncooperative power control game with exponential pricing for cognitive radio Network. *IEEE 65th Vehicular Technology Conference, 1,* 3125-3129.

Wang, W., Peng, T., & Wang, W. (2007). Optimal power control under interference temperature constraints in cognitive radio network. *IEEE Wireless Communications and Networking Conference, 1,* 116-120.

Wild, B., & Ramchandran, K. (2005). Detecting primary receivers for cognitive adio applications.

1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1, 124-130.

Zhang, D., & Tian, Z. (2007). Adaptive games for agile spectrum access based on extended Kalman filtering. *IEEE Journal of Selected Topics in Signal Processing, 1*(1), 79-90.

Zhang, L., Liang, Y.-C., & Xin, Y. (2007). Joint admission control and power allocation for cognitive radio networks. *IEEE International Conference on Acoustics, Speech and Signal Processing, 3*, 673-676.

KEY TERMS

Cognitive Radio: A new concept of radio transceivers, which have the distinguishing characteristics of reconfigurability and cognition.

Primary Users: In licensed bandwidth environments, users with the right to use the spectrum.

Secondary Users: In licensed bandwidth environments, users with the right to opportunistically use the spectrum by means of spectrum holes.

Spectrum Holes: Frequency bands that are idle in a given instant of time and place and are available for opportunistic allocation to users.

xGN: Next Generation Networks, characterized by the hierarchical structure of the users, coexisting primary or licensed users and secondary or unlicensed users.

Wireless: It refers to any type of electrical or electronic operation which is accomplished without the use of a wired connection.

Chapter LI
Game Theory for PHY Layer and MAC Sublayer in Wireless Telecommunications

J. Joaquín Escudero-Garzás
University Carlos III de Madrid, Spain

Ana García-Armada
University Carlos III de Madrid, Spain

ABSTRACT

The aim of this chapter is to address the role of a novel concept in wireless telecommunications: Game theory. Game theory is a branch of applied mathematics, which has recently drawn attention as a powerful tool to solve complex problems in wireless environments. To fulfil the intended goal, this chapter introduces the most relevant concepts of game theory such as game, player, and strategy, and give an overview of the applications of game theory in wireless networks.

INTRODUCTION

In telecommunications, state-of-the-art trends are towards very sophisticated services the objective of which is to provide multimedia contents to fixed or mobile users. This implies a strict quality control to guarantee the delivery of these services to the clients, ensuring high quality of service (QoS) and large data rates. To meet such requirements, telecommunications resources are intensively used. Moreover, when providing multimedia services, their quality parameters and the available resources are interrelated, creating dependence relations between them. As one may perceive, the task of a joint management of all these variables with the aim of exploiting the network in an optimized manner requires the utilization of tools capable of tackling such complex situations, and this task frequently becomes highly intricate when layer-by-layer solutions are applied. Then, novel approaches are used to solve complex telecommunications problems. One of these novel approaches is game theory, which can efficiently deal with the aforementioned problem of joint management of resources, service provisioning and network exploitation.

From its beginning in the middle of 20th century game theory has been intensively applied to economics, evolutionary biology and psychology, among others. Its utilization in telecommunications started at the late 1990s for the resolution of problems related to networking (flow and congestion control, routing), but its most interesting applications came relatively recently: its potential utilization for wireless systems.

The objective of this chapter is to provide an overview of game theory for telecommunications. Once the basic concepts are introduced, most relevant and recent works will be described.

GAME THEORY FOR WIRELESS COMMUNICATIONS

Why may game theory be applied to wireless telecommunications? The answer to this question rests on the very nature of game theory. It formalizes the interaction among autonomous agents (players) with selfish and even opposite objectives, and defines what a solution to the stated problem may be. It is clear then that by means of game theory the design and configuration of wireless networks such as ad hoc, sensor and mesh networks may be addressed in order to cope with limitations as power constraint, decentralized operation, interference mitigation and efficient multi-hop routing.

Game-theoretic concepts have mostly been applied in wireless applications to design network layer, flow control mechanisms and routing algorithms. Game theory has also demonstrated its validity for power allocation problem in wireless networks, where the network (e.g. the base station in centralized configurations) allocates the available resource (power) to the players (transmitters) with different types of constraints, e.g. interference, energy minimization or minimum required bit-rate constraints. In a similar manner, bandwidth allocation problems may be solved for this type of networks considering constraints such as the mutual interference among the transmitters. An alternative approach to obtaining solutions in telecommunication networks, e.g. routing protocols, is the usage of pricing: pricing schemes charge each node or customer locally for the resources he has used. An overview of the most relevant works on these areas is developed in the next section.

PHYSICAL (PHY) LAYER and MEDIUM ACCESS CONTROL (MAC) SUBLAYER problems in wireless networks have not received too much attention up to recent dates in the context of game theory. For instance, in the case of PHY layer, we may consider cooperative diversity (explained in detail in a subsequent section) as a suitable technique to be studied applying game theory given the intrinsic selfish behaviour of the users of the network. With respect to the MAC sublayer, in addition to allocation applications, some game theoretic approaches have attempted to model packet radio protocols (Aloha, Slotted-Aloha) and channel assignment strategies. These applications demonstrate how promising game theory is for PHY and MAC LAYERS design.

Game theory, as we mentioned in the introduction, is a mathematical tool, but it is not very familiar to telecommunications. Hence, we will devote the remaining of this section to introduce the most relevant concepts and elements used in the literature up to now to address wireless systems by means of game theory. Although we consider the overview provided in this section adequate for the comprehension of the rest of this chapter, the reader may find it useful to complement this introduction to game theory with (MacKenzie and DaSilva, 2006).

What is a Game?

Let us start by defining the very concept of game. A game is (Osborne, 2004) “a description of strategic interaction that includes the constraints on the actions that the players can take and the players' interests, but does not specify the actions that the players do take". The ultimate aim of the game formalism is to provide a solution consisting of a set of the different players' outcomes. Those outcomes are the result of jointly maximizing each player's utility, by means of the assignment to each player of a certain quantity of resources (money, power), considered as variables in their benefit functions. Then, it is clear that the goal of the players is to obtain a benefit as high as possible.

To Optimize or to Play a Game?

Game theory is worth being used for different telecommunication problems, but it must be handled with care. When we are trying to solve a problem in which the aim is to achieve the maximum benefit, sometimes the best tool to be used is not game theory. Therefore, we must be completely sure that we have not an optimization problem for a single decision maker, rather than a situation that involves several decision makers that are trying to get as much individual benefit as possible.

To make this clear, let us see an example. Consider the situation of telecommunications monopoly market, where there is only a telecommunications services provider for the users: The company will try to maximize its benefit performing an optimization of its resources; the optimization is based on the interaction with the users. On the other hand, for an oligopoly market where other telcos have entered, pricing games among all the companies take place and they respond in this context to the decisions of the consumers.

The Basic Elements of a Game: Players, Strategies and Benefit

The aim of this section is to provide to the reader an introductory overview on game theory. Then, we must start by formally defining a *game* in its *strategic form*: a game G is defined by the tuple {N, $\{\Sigma_i\}$, $\{U_i\}$ } $i \in$ N, where N = {1, ..., |N|} represents the set of players, $\Sigma_i = \{\sigma_1, ... \sigma_k\}$ represents the set of *k* pure strategies available for player *i*, and U_i represents the benefit for player *i*. If $S = \{\xi_1, ..., \xi_{|N|}\}$, for all $\xi_i \in \Sigma_i$, represents the Cartesian product of the set of strategies for all the players, then $U_i : S \rightarrow T$, being T_A subset of $\mathfrak{R}$. Note that each N-tuple $(\xi_1, ..., \xi_{|N|})$ is a possible outcome of the game.

Another possibility of representing a game is the *extensive form*, where the game becomes a tree: the nodes of the tree stand for the different stages of the game, because of the moves of the players, and each branch represents a possible action that the player may take.

Players. As we have introduced at the beginning of this section, a game is made up of three fundamental components: the players, their strategies and their benefit. The *players* are the decision makers interacting in the scenario where they play the game. Their final goal is to optimize the benefit they receive. In wireless communications, players model very frequently the nodes of the network or the users, but depending on the specific situation, players may be different elements of a telecommunications network, such as the links between nodes, or even telecommunications operators.

Strategies. Players will have different choices or alternative actions to play the game, and the different choices for each player are known as his *strategies*; for instance, the strategies for a certain node in a wireless network could be the selection of a modulation and coding scheme, his transmit power level or bit-rate. Once all players have selected their strategies, this determines the outcome of the game for the player.

The strategies that will yield the outcome of the game may be stated in two manners: The players of a game may act in a certain way breaking down to a set of strategies, which can be distinguished as *pure* and *mixed* strategies whether the player's behaviour is deterministic or probabilistic respectively. Usually, the set of strategies is made finite $\Sigma_i = \{\sigma_1, \ldots \sigma_k\}$, and each σ_j is a *pure strategy*. On the other hand, a *mixed strategy* over the set of pure strategies is a vector $\mathbf{p} = \{p_1, \ldots p_k\}$ with nonnegative values which sum up to 1, and where each coordinate p_j of $\mathbf{p}$ represents the probability for a player to choose strategy σ_j

Benefit: payoff and utility. Players receive a benefit when they choose a strategy, always attempting to get as much benefit as possible, and conditioned to other players' preferences and strategies of IR. In other words, players play the game to optimize their benefit. A players' benefit is specified by a real value assigned to the specific actions or strategies chosen by the players. Obviously, the larger this value the higher the player's benefit.

In obtaining the benefit, there exist two possibilities: it may be defined by means of a finite set of values or *payoffs*, so in this case the benefit is defined by $Y_i : \Sigma \to T$, being T a finite subset of T. The second possibility is by defining a *utility function*, which associates a number to the chosen strategy by means of a real function, that is, the benefit is expressed as $Y_i : \Sigma \to \Re$. As an example, a typical utility function in wireless networks is the bit-rate (bits per second) per consumed power (Watts) achieved in a particular interference situation.

Playing a Communications Game: Cooperation and Equilibrium

When playing a game, players may take their decisions knowing other players' set of strategies but without any knowledge of other players' chosen actions. In this case, the game is *non-cooperative*: the players do not exchange any information with each other to achieve the maximum selfish benefit.

An issue of paramount importance on a game is that the game must provide an outcome. For non-cooperative games, a stable outcome corresponds to a *NASH EQUILIBRIUM* (Nash, 1951), a concept directly related with utility functions. A set of strategies for each player $\mathbf{s^*} = \{\sigma_1^*, \ldots \sigma_N^*\}$ is a Nash equilibrium of $\Gamma = \{N, \{\Sigma_i\}, \{Y_i\}\}$ if no player can unilaterally improve his own utility while the rest maintaining their strategies fixed; mathematically, for every *j* player holds $Y_j(\sigma_j^*, \sigma_{-j}^*) \geq Y_j(\sigma_j, \sigma_{-j}^*)$ for all $\sigma_j \in \Sigma_j$, where σ_{-j}^* denotes de vector of strategies for all players except player *j*.

NASH EQUILIBRIUM may also be interpreted from a second viewpoint: $\mathbf{s^*} = \{\sigma_1^*, \ldots \sigma_N^*\}$ is a Nash equilibrium of $\Gamma = \{N, \{\Sigma_i\}, \{Y_i\}\}$ if σ_j^* is the best response (strategy) to σ_{-j}^*.

Players may move simultaneously, or sequentially. When they play simultaneously, most of the times they have no knowledge about what the other players do, and these games are called *static games*. If players move sequentially, they might have some information about other players' moves, or they might know everything; in the former case, we are talking about *dynamic games of partial information*, and in the latter we will refer to them as *dynamic games of perfect* or *complete information*. Nevertheless, sometimes, to achieve a solution to the game the players must interact several times, knowing the history of their previous movements and outcomes, and this interaction is modelled using *repeated games*. The utilization of distributed algorithms with partial information, referred sometimes in the literature as dynamic games, leads to tractable algorithms for players with processing constraints. The result is also a Nash equilibrium, since these algorithms converge to this point.

We have seen the importance of NASH EQUILIBRIUM when playing games. Therefore, proving the uniqueness and existence of NASH EQUILIBRIUM is a must in practical wireless communications games. Moreover, the NASH EQUILIBRIUM, even existing and being unique, is not optimal in general for the purpose of the communications system. On the contrary, there are situations in which players may interact before playing the game and jointly agree their actions. In this case, we say a game is *cooperative* and the players can obtain optimal outcomes, namely *Pareto optimal*. For a cooperative game, the solution of the game **s*** is Pareto-optimal if there exists no other solution for which one or more players could improve their utilities without reducing the utility of any of the other players. Therefore, any other *j*th player's strategy σ_j different from σ_j^* will cause a decrease in, at least, the utility of one of the other players, i.e. there exists at least one *j* player such that $Y_j(\sigma_j^*, \sigma_{-j}^*) \geq Y_j(\sigma_j, \sigma_{-j})$.

In the context of game theory for wireless communications, it is widely accepted that non-cooperative game-theoretic approaches are more suitable for distributed systems, given the additional complexity intrinsic to cooperative games. On the other hand, centralized networks are commonly modelled by cooperative games, since the central node or sink asks for information to the rest of nodes of the network. We detail some examples of both cases in the next section. However, as the reader may guess, these rules are not fixed: for example, cooperative games can also be applied to distributed network problems, as we shall show by means of the use of coalitional games.

Hitherto, we have considered Nash equilibria as potential solutions for a game. However, it might be appropriate for the system to model a game with the so-called *Stackelberg equilibrium* to define a different type of solution. When the solution comes from a Stackelberg equilibrium, the moves of the users shall follow the move of the network manager, which is acting as a "leader".

Specific Types of Games

Depending on a series of details that characterize the game, in the literature we can identify some specific game models when applying game theory to wireless communications. Among these models, potential and supemodular games are particularly interesting: the existence of Nash equilibrium is not guaranteed for games in general, while Nash equilibrium always exists in these two particular types of games; furthermore, the game converges to this Nash equilibrium. In the following, we shall briefly introduce these types of games and provide some examples of application.

A *potential game* is a game characterized in its strategic form by a function $\varsigma: \Sigma \rightarrow Y$, such that if a unilateral deviation occurs, the change in ς, namely $\Delta\varsigma$, is reflected in a change of the utility function seen by the deviating player *i* ΔY_i. In other words, a potential function exists and reflects exactly any unilateral change in the utility function of any player. The rationale for the use of potential games, as we have mentioned in the above paragraph, is that they converge to a NASH EQUILIBRIUM. As examples when applying potential games, we highlight the following references: Nie and Comaniciu (2006) show the application of potential games for channel allocation in wireless networks; Scutari, Barbarossa and Palomar (2006) yield a framework for power allocation using potential games; and in (Hicks, Mackenzie, Neel and Reed, 2004), interference avoidance is addressed also in the context of potential games.

A game $\Gamma = \{N, \Sigma, \{Y_i\}$ is a *supermodular game* if for each player $i \in N$, the strategy space Σ_i is a nonempty and compact subset of $\Re$, the utility functions Y_i are continuous and has increasing differences between player *i*'s strategy and any other player's strategy. Although the formulation is complicated, supermodular games posses the interesting property that there always exists a NASH EQUILIBRIUM. For this reason, this type of games has been extensively used to solve power management in wireless networks; to illustrate this we refer the reader to the works of Saraydar, Mandayam and Goodman (2002) and Altman and Altman (2003) as examples.

FUNDAMENTAL RESULTS IN GAME THEORY FOR WIRELESS COMMNICATIONS

In this section, we describe the relevant issues to which game theory is applied in wireless communications; for example, power allocation, the backoff attack problem or cooperative diversity. For these problems, we also summarize the most highlighted works and results that the usage of game theory has provided to date. Of course, game theory has also come up with solutions others than for the applications we mention, e. g. flow (Altman and Basar, 1998) and congestion (Hwang, Tan, Hsiao and Wu, 2005) control, quality of service (QoS) (Wang and Ramanathan, 2005) or routing (Altman, Boulogne, El-Azouzi, Jimenez and Wynter, 2006). Nevertheless, our goal is to provide a vast overview on game theory -based wireless communications for PHY and MAC LAYER. In detail, we shall analyze power management and interference avoidance in the context of PHY layer, the backoff attack problem and random access protocols for MAC sublayer, and three issues related to both PHY and MAC: spectrum management, cooperative diversity and cross-layer design.

PHYSICAL LAYER

Power Control and Management

Power-based design is one of the crucial wireless communications subjects, given the need of a tradeoff between the transmission power and the potential interference caused: the higher the transmit power to achieve large bit-rate, the higher the interference it may cause. Game theory has emerged, from its earliest utilization in communications, as a very adequate tool when power assignment becomes a problem if the aim is to satisfy users' requirements fairly. Moreover, limitation of energy of battery-powered devices in wireless networks such as sensor and ad hoc networks has arisen in the last few years as the most restricting factor of performance in these networks. To address the energy-efficient use of batteries, power must be treated as a scarce resource in the context of game theory. In the following, we will point up the most recent research in power management, since references corresponding to the early works in this field are detailed in (Altamn et al., 2006).

A major concern on power management is the level of interference a transmitter may cause over other users, especially when we face some modern wireless transmission technologies, of which MIMO (Multiple Input Multiple Output) systems and ultra-wide band (UWB) are illustrative examples. To solve the problem of interference using Game Theory, one of the two following approaches usually applies: the harmonization of selfish interests among users desiring to transmit at maximum power or directly penalizing (charging) those users employing their power unfairly. In impulse-radio (IR) UWB systems, the particular effect of self-interference must be considered and channel fading cannot be assumed flat. These systems are suitable to be analyzed from a game-theoretic perspective as done in (Bacci, Luise, Poor and Tulino, 2007), by means of a non-cooperative power control game. For interference mitigation in ad hoc MIMO systems and additional optimization of the consumption of batteries, Liang and Dandekar (2007) propose a power control game. They formulate the problem from an information-theoretic point of view: minimizing interference by using power control increases MIMO capacity and lifetime of batteries. The remarkable point in this approach is the fact that the players are the links, which choose a transmit power and want to maximize their utilities, consisting on the capacity of each link diminished by a second term proportional to the power used in this link.

The purpose of power control for uplink of Code Division Multiple Access (CDMA) is to assure for each user a given QoS without interfering with the other users. This topic has been addressed in detail using Game Theory, and a suitable metric of satisfaction for the users is the resulting signal-to-interference-ratio (SIR) in the uplink. For example, Alpcan, Basar and Dey (2006) analyze the case of a multicell CDMA network where the utility of a player grows as the outage probability decreases and the cost function reflects the battery usage.

As regards energy constraints, efficient power control has been tackled frequently through pricing, an adequate framework to deal with selfish behaviour of players (wireless terminals in our case) whose actions would affect to other players' utilities. By using pricing, the objective is twofold: generating revenue for the whole system, and encouraging players to use system resources efficiently. The desired result through pricing has been largely obtained by means of utility functions of the form of throughput per energy (bits per Joule) or throughput per power (bits per Watt), since the utility represents the benefit obtained for one unit of consumed resources (see again (Saraydar et al., 2002)).

Interference Avoidance

Interference is considered as a limitation factor of performance in the design of wireless systems, so the interest on its avoidance justifies the numerous literature in this matter. Clearly, interference analysis is closely related to power management, and they are often jointly addressed as we have seen in the precedent subsection. Let us see an example where power control games are not the only method to minimize interference in a game-theoretic framework. The Lacatus and Popescu's (2007) proposal is a distributed adaptive algorithm for CDMA systems to obtain optimal codewords and power allocation for a required signal-to interference plus noise ratio (SINR); the game is formulated for joint codeword and power adaptation and separated into two non-cooperative subgames: a codeword adaptation subgame and a power adaptation subgame. Sung, Shum and Leung (2006) introduced the concept of *subgame*: when the utility function of the game is separable with respect to the variables that define the users' strategy the game may be separated into as many independent subgames as variables define the strategy; and each subgame utility function will depend exclusively on one of the variables.

MAC SUBLAYER

The Backoff Attack Problem

One of the security problems recently found in networks based on IEEE 802.11 standard has to do with Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA), the contention mechanism of MAC sublayer Distributed Coordination Function (DCF) defined in the standard. The vulnerability, known in the literature as *backoff attack*, is related to the *contention window* (CW), which sets a countdown to transmit each time the channel is sensed as idle. Then the user will transmit with a certain probability when the countdown is over. The problem is that each user is able to set the CW locally, thus by giving a small value to the CW the user will have more opportunities to transmit: the lower the CW, the higher the throughput. This is, of course, a misbehaving procedure and those users are acting as *cheaters*.

In order to solve the backoff attack problem, different authors have made game-theoretic approaches, and the key point is to design a game in which the cheater is discouraged on malicious behaviour. Among all works that address backoff attack, the perspective taken by Konorski (2006) is particularly interesting. He proposes a CSMA/CA repeated game (CRISP, *Cooperation via Randomized Inclination to Selfish/Greedy Play*), based on classical Prisoner's Dilemma (Binmore, 2007). He proposes the utility function of user *k* is directly related to his previous bandwidth allocation b_k and this bandwidth will be dramatically inferior if a user acts cheating about his CW instead of acting honestly, e.g., playing CRISP; consequently, users are discouraged on malicious behaviour. Equally interesting is the approach deployed by Cagalj, Ganeriwal, Aad and Hubaux (2005): initially, all the nodes are considered cheaters and they have the ability to measure the others' throughput, and so detecting which nodes are cheating, e.g. not cooperating, and thus transmitting at higher rates than tolerated. Of course, the game is cooperative and the nodes share the necessary information to detect whether a user is cheating or not.

Random Access Protocols

In wireless systems with random access, one of the most popular protocols is CSMA. In CSMA systems, user's access to wireless channel is probabilistic, and is characterized by the *persistence probability*, defined as the probability which the user will attempt to transmit with, once the channel is idle. Researchers have studied the persistence probability in the framework of Game Theory and they have mainly focused their works on calculating the persistence probabilities to optimize their transmission, either by maximizing the throughput or by minimizing the probability of collision. The key question for a user is to reach equilibrium: neither to be too aggressive, with high values of the persistence probability, nor to be too conservative. For the former case, the number of collisions would increase and for the latter, the user would miss opportunities of transmission; in both cases, the performance deteriorates.

The following two examples make clear how to apply game theory to CSMA. Altman, Borkar and Kherani (2004) optimize the persistence probability for CSMA networks in this direction. The key point of the non-cooperative game is the combination of transmission and reception probabilities through the utility function of the nodes, that is, both roles of a node, as transmitter and as receiver, are considered. Another option is by designing a MAC protocol which makes use of links as players of the game (Tang, Lee, Huang, Chiang and Calderbank, 2006), and the utility of each link will depend on his own transmission and transmissions of other links. Each link's strategy is the persistence probability of the random access, and the utility function has the form of the expected reward (depending on the probability of successful transmission) minus the expected cost (depending on the probability of transmission failure).

By its simplicity for implementation, Aloha protocols have been very popular in packet radio networks. The first and most relevant attempt of addressing Aloha protocol using game theory was provided by Wu and Wang (1995). They analyzed Slotted Aloha in a game-theoretic framework where users have different QoS requirements. The objective of the game is to achieve fairness among users when accessing the channel, trying not to favour those nodes closer to the access point. The strategies are made up by the pair packet retransmission probability (that is analogous to persistence probability of CSMA) and power transmission, resulting in a multi-objective approach, which leads to the optimization of throughput and average delay.

A problem that has gained more relevance with the appearance of coexisting wireless technologies, as WiFi and WiMAX, is the presence of several access points (AP) in the same service area, based or not on the same technology. The solution to medium access in this competitive environment provides a good opportunity for game theory, given the natural selfish interest of each AP. Although involving not only MAC aspects in a sort of cross-layer protocol, a session admission and rate control (ARC) protocol with multiple access providers in a competitive market is considered in (Lin, Chatterjee, Das and Basu, 2005). Being competitive, the market allows performing *charming* to users, that is, they have the freedom of switching to a different provider. The viewpoint of providers is adopted when designing the ARC algorithm, since the objective is to maximize the provider's revenue. For session admission, a non-cooperative game is played among each provider and his customers: the provider tries to maximize his revenue, and users want to maximize their payoffs that include a penalization in the case they decide to churn unilaterally before the end of the contract. Users are differentiated into classes according to their packet blocking rate. The rate control scheme is based on allocating power to the users: if a new admission occurs the transmission rate of users in the same class as well as in lower classes is reduced.

SPECTRUM MANAGEMENT AND SPECTRUM SHARING

Given the well-known scarcity of spectrum for wireless communications, the research community has done many efforts to use it efficiently and the utilization of game theory has played a relevant role in these studies. One of the approaches deals with integration of 802.11 WLAN with 802.16 multihop mesh networks, to relay WLAN traffic to Internet; we can cite, for instance, Niyato and Hosssain (2007). These authors define the following framework to analyze the coexistence of wireless networks: a connection is admitted if and only if the global utility is improved. If the network admits the connection, all the connections play a bargaining game, being the solution a Pareto-optimal assignment of the bandwidth to the established connections.

In centralized networks, a common approach is to consider *mechanisms*, a mathematical tool well studied in Economics, where the users communicate their preferences to a base station or a similar entity. We can see a *mechanism* as a three-stage process. First, a leader announces the rules (named here "outcomes") of the mechanism. At a second stage, the agents communicate to the leader, by means of messages, their private information. Finally, the leader allocates resources according to the rules and the private information provided by the agents. Price and Javidi (2007) propose a rate allocation mechanism for the downlink of single-hop networks. The ultimate aim of the algorithm is the design of a centralized scheduler considering the efficient use of uplink for the allocation of rate to the downlink. Proceeding in such a way, the users in the uplink are forced to reveal their utility truthfully, intended as private information, and they are not tempted to cheat about their needed rate, since users are charged in terms of downlink for their consumption of uplink.

A different sight is to consider the following game: the players are networks deployed in the same geographical area, sharing the available channels. Wendorf and Seidenberg (2007) follow this approach, by proposing a methodology for channel assignment in spectrum-agile scenarios, where networks can dynamically switch their communication channels. These authors propose a game, which they refer to as "channel-change game", with two networks as players. The networks can switch channel or keep it when they are using the same channel. If a network switches to a different channel, it incurs in a delay cost, but if both remain in the same channel, there is a blocking. The solution of the game leads to the use of mixed strategies, since pure strategies yield to a blocking situation.

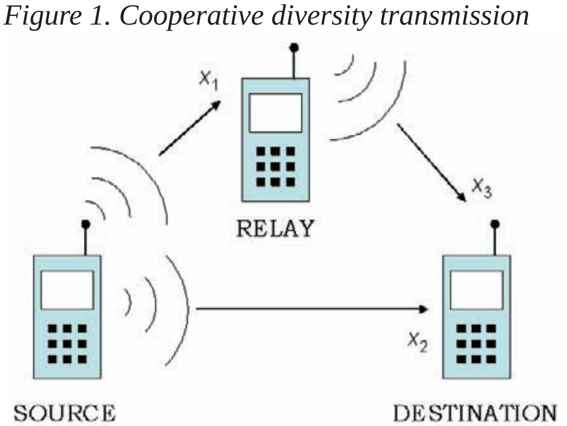
COOPERATIVE DIVERSITY

In large wireless networks, cooperation among the nodes is essential to achieve end-to-end communications when the source and the destination nodes are out of range, i.e. they cannot establish a direct link. In packet networks, end-to-end communications consist of multihop communications, and some of the intermediate nodes between the source and the destination act as *relays*, carrying out packet forwarding or signal forwarding. Novel techniques in wireless communications rely on cooperation among the users for the transmission of other users' information.

Cooperative diversity allows wireless nodes to collaborate with each other to complete their end-to-end transmissions in a multi-hop network: the source node sends the signal directly to the destination, and the same signal received (x_1) by an intermediate node (relay) is retransmitted to the destination. The signal obtained in the receiver is a combination of both direct x_2 and retransmitted received x_3 signals (see *Figure 1*). The crucial difference then between cooperative diversity and simple relaying is the spatial diversity, due to the existence of two paths that conveys signals to the destination. Due to their implicit use of cooperation, the closeness between both cooperative diversity and relaying schemes is clear.. Consequently, in this subsection we will also refer to relay approaches based on Game Theory.

As pointed out in the precedent paragraph, cooperative diversity exploits spatial diversity and the result is the improvement of the reliability of the communication in terms of, for example, bit error rate. Naturally, the first intuition is to think of cooperation from an selfish viewpoint, since a cooperative node is spending his own resources (power, bit rate) for the transmission of other's information; nevertheless, the selfish and at the same time intelligent position is to cooperate as demonstrated in (Sendonaris, Erkip and Aazhang, 2003).

Figure 1. Cooperative diversity transmission



Cooperation in cooperative diversity goes further than the simple relay action extensively used for packet forwarding, as it is explained with the following different cooperation schemes. The simplest approach is known as *Amplify-and-Forward* (AF) scheme, in which the intermediate node acts as a relay, amplifying the received signal in accordance with its own power constraints. *Decode-and-Forward* (DF) schemes are more sophisticated: the intermediate node performs detection and decoding for the received signal, and encodes this signal again, therefore reducing the impact of the noise. While AF and DF schemes are fixed, *Selection-and-dynamic* schemes allows the cooperating terminals to adapt their transmissions according to the channel state.

Among cooperative diversity schemes, AF scheme is widely utilized since it is a very straightforward way of cooperation. Wang, Han and Liu (2006) formulate AF as a Stackelberg game, where the leader is the source node and the potential relays act as the followers. In the game, the source wants to buy power to the relays for retransmission, and the relays compete among themselves to provide the leader (i.e., the source) the required power at a suitable price. Also interesting is the study and application of AF and DF schemes in MIMO cellular environments, as shown in (Agustín, Muñoz and Vidal, 2004).

Stimulation of AF cooperative diversity for ad hoc networks is devised as a game with pricing according to Shastry and Adve (2006). The aim is to induce forwarding, ensuring that the users and the access point (AP) benefit from cooperation: users reimburse each other for forwarding and AP charges users for transmitting data packets. Users' preferences are to obtain as high throughput and low energy consumption as possible, so their utilities consist of the throughput per consumed energy (bits per second/joule). The game is played by two users with the possibility to cooperate in the form of a strategic game, with the goal to reach the Nash equilibrium, and therefore determining the power needed for cooperation. The role of the AP is to communicate to both users the reimbursement prices it has previously obtained by maximizing its own service price.

As Scutari, Barbarossa and Ludovici (2003) show, the design of optimal coding strategies for multihop networks alleviates the necessary signalling required for cooperative transmission by using cooperative diversity. These authors derive the optimal coding matrix by proposing a non-cooperative game, with the objective of maximizing the information rate between transmitter and receiver in the presence of interference. In this game, the players identify the links of the network, the transmit power for each user is his strategy, and the payoff corresponds to the information rate.

When using coalitional games, players are divided into "coalitions". If cooperation is formulated in terms of a coalition game, a coalition is a group of cooperative nodes (Mathur, Sankaranarayanan and Mandayam, 2006), and the nodes of a network may take part of the coalitions. The objective of the game proposed by Mathur et al. is to maximize the receiver's rate, based on the mutual information between the transmitters and the receivers making up the coalition. A node may decide to leave the coalition to achieve a higher rate.

Nodes acting as relays may be interested in cooperation, or may not. Srinivasan, Nuggehalli, Chiasserini and Rao (2005) address how to decide about cooperation by means of a non-cooperative game where the nodes are organized into classes depending on their energy constraints and lifetime expectation. In this game, the utility function is the number of relay requests generated by a node. The nodes can use Generous Tit-For-Tat (GTFT) strategies, a variation of TFT strategies: in a dynamic game, a node playing TFT mimics the action other player has carried out in the previous realization of the game; this would lead to a non-relaying action for the node playing TFT if the imitated node did not cooperate. When the nodes use GTFT they may act occasionally "generously" not mimicking the non-cooperative strategy. The implementation of GTFT entails a record of the past experience of each node. One of the advantages of the proposed game is that the solution is a Nash equilibrium that converges to a Pareto-optimal assignment of the number of connections to be relayed by a node.

Felegyhazi, et al. (2005) address cooperation for packet relaying in wireless sensor networks (WSN). The goal of the game is to determine if the controllers (here referred to as authorities) pertaining to different WSNs can help each other in forwarding packets and, at the same time, increase the lifetime of their own network, since the game assumes a cost of energy for the transmission, processing and reception of a packet. Also in the context of WSN, Crosby and Pissinou (2007) propose an algorithm for cooperation in wireless sensor networks. The main difference with the game of Felegyhazi et al.'s game is that the cooperation for packet forwarding is established among classes: sensor nodes are divided into classes according to a localization criterion; therefore, the nodes pertaining to the same class perform total cooperation. Crosby and Pissinou's algorithm is based on a type of strategy known as Grim Strategy. When using this strategy the players act as follows: as long as a player *A* cooperates the player *B* will cooperate, but if not, player *A* will not receive any cooperation until he shows his willingness of cooperating during a certain period of time. Crosby et al. define the Patient Grim Strategy, an adaptation of the Patient Strategy, which enforces cooperation by means of punishment: if a node rejects cooperation for *n* times, the non-cooperating node will abandon the network forever.

The idea of enforcing cooperation, mentioned in the above paragraph, is extensible to WLAN and WWAN networks. Wei and Gitlin (2005) propose a novel approach based on scheduling schemes for 3G networks. The schemes they propose encourage cooperation for relaying, analyzing them in a game-theoretic framework. They formulate a repeated game with a punishment mechanism that avoids the pure selfish strategy: if a node does not cooperate for relaying, he will be excluded from the network and he will be admitted again after cooperating a certain period of time.

CROSS-LAYER DESIGN

The classical strict layering approach to solve complex problems in engineering has a significant limitation since solving a certain problem in a given layer may produce degradation in adjacent layers, especially when it comes to wireless communications. These networks have a dynamic topology and their links exhibit a time-varying quality due to fading, shadowing and multiuser interference. If the usage of spectrum, energy or other radio resources is at a premium, we could expect a more efficient management of them if some cooperation exists between the upper (namely MAC and NET) and PHY layers.

Some significant studies have emerged recently making use of game theory on cross-layer design due to its inherent capability of optimization, and addressing different types of networks by exchanging cross-layer information. Ngo and Krishnamurthy (2007) provide a remarkable example of game-theoretic cross-layer design. In multipacket reception networks the Generalized Multi-Packet Reception (G-MPR) model is widely used, providing explicit incorporation of PHY layer channel state information (CSI) into the reception of MAC packets. Hence, they define a mapping of channel states to transmission probabilities known as transmission policy. The formulation of a non-cooperative game, where the players are the transmitting nodes and the strategies are the transmission policies, provides Nash equilibrium solutions if every node adopts transmission policies with the SINR threshold reception model.

FUTURE TRENDS

We have shown that the use of game theory to solve certain problems in wireless communications, such as power control, has an important background. Even so, some specific types of games such as supermodular and potential games may help solving problems in a more precise way. Moreover, most of the used games are non-cooperative; cooperative games, on the other side, provide optimal solutions in the form of Pareto optima, but at the expense of an exchange of information that may not be easy. Thus, for some problems, it would be worth to evaluate firstly if the game can be turned into a cooperative game in order to obtain the maximum benefit. Another powerful but at the same time complex tool are mechanisms, which were also introduced in section "Spectrum Management and Spectrum Sharing"; although the application of mechanisms for very specific situations is quite recent, we may envisage their use in decision-making problems. Moreover, they may result particularly useful when a set of rules must be fulfilled; in this case, the game framework is does not apply.

Cross-layer design approaches based on game theory are now emerging to deal with situations where a highly demanded resource affects the performance of other layers; in these cases, it becomes more effective to establish interlayer connections. Given that game theory is very suitable for global optimization, it appears as a very promising instrument for the future in this area.

The environment of wireless communications is getting more and more complex, due to the coexistence of different types of networks and standards. Let us take the example of cellular and ad hoc networks: for the former, a regular pattern and frequency allocation are considered in the deployment of the base stations infrastructure, and for ad hoc networks there is no planned deployment at all. Moreover, consider diverse WLAN and WWAN standards (IEEE 802.11/802.15/802.11/802.22): for each of them, the spectrum sharing and spectrum access strategies are different. The goal of designing MAC protocols in such unfriendly situation turns out to be a hard task and game theory has shown to be a suitable choice for sharing access in more simple scenarios; so it is time to verify its effectiveness for this challenge.

There exists a trend towards distributed organization in wireless systems such as ad hoc networks and Wireless Personal Area Networks (WPAN). Keeping in mind the distribution of tasks among the nodes of these networks, the most reasonable approach is to consider every node as a decision-maker that will interact with other network components. Consequently, game theory can provide the appropriate framework to address new challenges for networks based on distributed topologies, such as cooperative diversity.

CONCLUSION

Game theory is an interesting field of mathematics by itself, but becomes even more attractive when applied to engineering problems such as wireless communications. We have exposed how game theory has gained popularity as a solution method for power management, interference avoidance, the backoff attack problem, medium access and cooperative diversity. We have given references to where to find the most relevant work in areas out of the scope of this chapter, such as routing and flow control. We have shown that game theory is a useful tool to tackle wireless communications in which interdependent actions may condition the performance of both the individuals and the whole network.

ACKNOWLEDGMENT

The authors would like to thank Professor Bousoño-Calzón for his valuable comments and suggestions on Game Theory.

REFERENCES

Agustín, A., Muñoz, O., & Vidal, J. (2004). A game theoretic approach for cooperative MIMO schemes with cellular reuse of the relay slot. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 4*, 581-584.

Altman, E., & Altman, Z. (2003). S-modular games and power control in wireless networks. *IEEE Transactions on Automatic Control, 48*(5), 839-842.

Altman, E., & Basar, T. (1998). Multiuser rate-based flow control. *IEEE Transactions on Communications, 46*(7), 940-949.

Altman, E., Borkar, V. S., Kherani, A. A. (2004). Optimal random access in networks with two-way traffic. *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1*, 609-613.

Altman, E., Boulogne, T., El-Azouzi, R., Jiménez, T., & Wynter, L. (2006). A survey on networking games in telecommunications. *Elsevier Computers & Operations Research, 33*(2), 286-311.

Alpcan, T., Basar, T., & Dey, S. (2006). A power control game based on outage probabilities for multicell wireless data networks. *IEEE Transactions on Wireless Communications, 5*(4), 890-899.

Bacci, G., Luise, M., Poor, H. V., & Tulino, A. M. (2007). Energy Efficient Power Control in Impulse Radio UWB Wireless Networks. *IEEE Journal of Selected Topics in Signal Processing, 1*(3), 508-520.

Binmore, K. (2007). *Playing for Real. A Text on Game Theory*. New York, NY: Oxford University Press.

Cagalj, M., Ganeriwal, S., Aad, I., & Hubaux, J.-P. (2005). On selfish behaviour in CSMA/CA networks. *24th Annual Joint Conference of the IEEE Computer and Communications Societies, 4*, 2513-2524.

Crosby, G. V., & Pissinou, N. (2007). Evolution of cooperation in multi-class wireless sensor networks. *32nd IEEE Conference on Local Computer Networks, 1*, 489-495.

Felegyhazi, M., Hubaux, J.-P., & Buttyan, L. (2005). Cooperative packet forwarding in multi-domain sensor networks. *3rd IEEE International Conference on Pervasive Computing and Communications Workshops, 1*, 345-349.

Hicks, J. E., Mackenzie, A. B., Neel, J. A., & Reed, J. H. (2004). A game theory perspective on interference avoidance. *IEEE Global Telecommunications Conference, 1*, 257-261.

Hwang, K.-S., Tan, S.-W., Hsiao, M.-C., & Wu, C.-S. (2005). Cooperative multiagent congestion control for high-speed networks. *IEEE Transactions on Systems, Man and Cybernetics, Part B, 35*(2), 255-268.

Konorski, J. (2006). A game-theoretic study of CSMA/CA under a backoff attack. *IEEE/ACM Transactions on Networking, 14*(6), 1167-1178.

Lacatus, C., & Popescu, D. C. (2007). Adaptive Interference Avoidance for Dynamic Wireless Systems: A Game Theoretic Approach. *IEEE Journal of Selected Topics in Signal Processing, 1*(1), 189-202.

Liang, C., & Dandekar, K. R. (2007). Power Management in MIMO Ad Hoc Networks: A Game-Theoretic Approach. *IEEE Transactions on Wireless Communications, 6*(4), 1164-1170.

Lin, H., Chatterjee, M., Das, S. K., & Basu, K. (2005). ARC: an integrated admission and rate control framework for competitive wireless CDMA data networks using noncooperative games. *IEEE Transactions on Mobile Computing, 4*(3), 243-258.

MacKenzie, A., & DaSilva, L. (2006). *Game Theory for Wireless Engineers*. Ft. Collins, CO: Morgan & Claypool.

Mathur, S., Sankaranarayanan, L. & Mandayam, N.B. (2006). Coalitional games in receiver cooperation for spectrum sharing. 40th *Annual Conference on Information Sciences and Systems, 1*, 949-954).

Nash, J. (1951). Non-cooperative Games. *Annals of Mathematics, 5*(2), 286-295.

Ngo, M. H., & Krishnamurthy, V. (2007). Game theoretic cross-layer transmission policies in multipacket reception wireless networks. *IEEE Transactions Signal Processing, 55*(5-1), 1911-1926.

Nie, N., & Comaniciu, C. (2006). Adaptive channel allocation spectrum etiquette for cognitive radio networks. *Springer Mobile Networks and Applications, 11*(6), 779-797.

Niyato, D., & Hossain, E. (2007). Integration of IEEE 802.11 WLANs with IEEE 802.16-based multihop infrastructure mesh/relay networks: A game-theoretic approach to radio resource management. *IEEE Network, 21*(3), 6-14.

Osborne, M. J. (2004). *An introduction to Game Theory*. New York, NY: Oxford University Press.

Price, J., & Javidi, T. (2007). Leveraging downlink for efficient uplink allocation in a single-hop wireless network. *IEEE Transactions on Information Theory, 53*(11), 4330-4339.

Saraydar, C. U., Mandayam, N. B., & Goodman, D. J. (2002). Efficient power control via pricing in wireless data networks. *IEEE Transactions on Communications, 50*(2), 291-303.

Scutari, G., Barbarossa, S., & Ludovici, D. (2003). Cooperation diversity in multihop wireless networks using opportunistic driven multiple access. *4th IEEE Workshop on Signal Processing Advances in Wireless Communications, 1*, 170-174.

Scutari, G., Barbarossa, S., & Palomar, D. P. (2006). Potential games: a framework for vector power control problems with coupled constraints. *IEEE International Conference on Acoustics, Speech and Signal Processing, 4*, 241-244.

Sendonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity. Part I. System description. *IEEE Transactions on Communications, 5*(11), 1927-1938.

Shastry, N., & Adve, R.S. (2006). Stimulating cooperative diversity in wireless ad hoc networks through pricing. *IEEE International Conference on Communications, 8*, 3747-3752).

Srinivasan, V., Nuggehalli, P., Chiasserini, C.-F., & Rao, R. R. (2005). An analytical approach to the study of cooperation in wireless ad hoc networks. *IEEE Transactions on Wireless Communications,* 4 (2), 722-733.

St. Jean, C. A., & Jabbari, B. (2004). Bayesian game-theoretic modeling of transmit power determination in a self-organizing CDMA wireless network. *IEEE 60th Vehicular Technology Conference, 5*, 3496-3500.

Tang, A., Lee, J.-W., Huang, J., Chiang, M., & Calderbank, A. R. (2006). Reverse engineering MAC. *4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 1*, 1-11.

Wang, B., Han, Z., & Liu, K. J. R.(2006). Stackelberg game for distributed resource allocation over multiuser cooperative communications networks. *IEEE Global Communications Conference,1*, 1-5.

Wang, K.-C., & Ramanathan, P. (2005). QoS assurances through class selection and proportional differentiation in wireless networks. *IEEE Journal on Selected Areas in Communications, 23*(3), 573-584.

Wei, H.-Y., & Gitlin, R. D. (2005). Incentive scheduling for cooperative relay in WWAN/WLAN two-hop-relay network. *IEEE Wireless Communications and Networking Conference, 3*, 1696-1701.

Wendorf, R. G., & Seidenberg, H. B. (2007). Channel-change games for highly interfering spectrum-agile wireless networks. *2nd International Symposium on Wireless Pervasive Computing, 1*, 611-616.

Wu, B., & Wang, Q. (1995). A multiobjective analytic framework for slotted ALOHA wireless LANs. *4th IEEE International Conference on Universal Personal Communications, 1*, 12-16.

KEY TERMS

Game: Mathematical formulation where the players choose strategies that will maximize their benefit.

Game Theory: It is a branch of applied mathematics that is often used in the context of economics, which studies interactions between decision-makers.

Wireless: It refers to any type of electrical or electronic operation that is accomplished without the use of a wired connection.

Player: A decision-maker in the game-theoretic framework.

Strategy: The different choices or alternative actions a player has to play a game.

Utility Function: Function associated to each player of a game which maps the resource consumption into a number so as to represent player's preference.

Chapter LII
Agent Based Product Negotiation Models in Mobile Commerce

S. S. Manvi
REVA Institute of Technology and Management, India

L. B. Bhajantri
Basaveshwar Engineering College, India

ABSTRACT

This chapter brings out various issues in M-Commerce and describes various agent-based product negotiation models in mobile commerce environment. The negotiation models discussed are based on auctions, trade off, argumentation, contract Net, bilateral and game theory. Mobile commerce (M-Commerce) is an emerging manifestation of Internet electronic commerce that bridges the domains of Internet, mobile computing and wireless telecommunications in order to provide an array of sophisticated services to mobile users. The usages of M-Commerce services in commercial activities are expected to dominate the world. Commercial activities through wireless are gradually changing the daily practice and future possibilities. This opens many new possibilities, opportunities and challenges in mobile commerce. Using software agent technology by providing intelligence, autonomous, customized, adaptable and flexible services, can enhance M-Commerce activities. The agent technology supports Component Based Software Engineering (CBSE) thereby allowing software developers to create dynamic software architecture for m-commerce applications.

MOBILE COMMERCE

M-commerce (mobile commerce) is an evolving area of e-commerce, where users can interact with the service providers through a mobile and wireless network, using mobile devices (PDA's, mobile phones, laptops, etc.) for information retrieval and transaction processing. E-commerce (electronic commerce) refers to the field of marketing, buying, selling, distributing and servicing different products and/or services over the Internet. It aims at using electronic business applications for the purpose of commercial transactions. M-commerce is always reliable and easy to access as compared to e-commerce since wireless handheld devices are easy to carry anywhere because of miniature size and low weight.

The wide deployment of wireless networks, mobile technologies and the significant increase in the number of mobile device users have created a very strong demand for emerging mobile commerce applications and services. It opens new avenues for businesses. Many researchers assume wireless technology supporting m-commerce will change the global economic environment to a large extent.

Characteristics of M-Commerce

The dynamic development of mobile commerce and the increasing integration in the Internet have opened up an entirely new research and development field (George, 2002). Several frameworks have emerged to define the wide range of mobile computing applications, characteristics and services. One such framework presents classes of mobile applications ranging from retail, auction, financial services, advertising, and inventory management to mobile office, distance education, and data center. Another framework groups mobile services into goods, services, and content for consumer e-commerce, and activities among trading partners (Keng, 2001).

The m-commerce is characterized by mobility, ubiquity and localization (Shaffer, 2000, Ishara, 2003). Ubiquity is achieved since the users are mobile (carry cell phones or other mobile devices from place to place) and has the ability to receive information and perform transactions at any time and any location. The technologies like Global Positioning System (GPS) or Time of Arrival (TOA) enable m-commerce users to access information and services specific to their location. Mobile devices in an m-commerce environment are typically operated by and configured for a single user and thus services can be personalized.

M-Commerce Applications

Following are the kinds of applications provided in m-commerce: B2C (Business to consumer)/ C2C (Consumer to Consumer) m-transactions services such as stock trading, mobile banking and location based advertising. B2B (Business to Business) m-transactions such as access to B2B e-commerce marketplaces, access to bilateral online trading systems, and banking services. Personal life management of everyday activities, such as e-mail, chatting, instant messaging, entertainment, multipurpose remote controls, and information services such as weather forecasts and news updates. Mobile office support using e-mail, calendar, groupware and information services. Mobile operations for sales force support, field service management, fleet management, and remote monitoring.

M-Commerce customers may be more demanding and less enduring than e-commerce users (Irvine, 2001). Mobile users require value-added services that can be feasible or non-feasible according to existing technologies and m-commerce constraints such as, performance, reliability, security, ease of use, bandwidth, etc.

Following are the advantages of m-commerce. Customers are very easily accessible since most mobile device users walk around with their gadgets; hence sellers can communicate with the buyers immediately. Commerce activities' using mobile devices increases customer satisfaction, cost savings; offers new business opportunities. Commerce services can be personalized and the user can have control over the transactions. M-commerce enhances customer and seller relationships since communication is ubiquitous.

The following sections describe the issues in m- commerce, negotiation in m-commerce, role of agents in m-commerce and agent based negotiation models. Finally, the chapter is summarized along with some of the key terms.

ISSUES IN MOBILE COMMERCE

This emerging area of m-commerce has several issues and because of new technology, novel applications, and increased pervasiveness (Tarasewich, 2002; Schwiderski, 2000). The issues are listed as follows: device type, limited input/output interface, processing capacity, low bandwidth, frequent disconnections, mobility, geographic location, device and personal identity, security, business, product and QoS negotiation, etc. *Type of devices*: Various types of hand held devices by manufacturers are available which necessitates compatibility among the devices. *Limited input/output interface* (screens, keyboards, etc): the small screens and keypads of mobile devices limit the size of information that can be displayed on the mobile screen and make data entry difficult and uncomfortable especially for longer messages and browsing of information. *Processing capacity*: limited processing power with small memory chip and limited power supply. *Low bandwidth*: there are still several constraints for transmitting audio and video over mobile devices. Although they are continuously increasing, transmission rates are still very low for wireless devices compared to fixed-wired networks. *Frequent disconnections*: cell interference, limited cell capacity or lack of network coverage can lead to frequent disconnections. *Mobility*: the user's move from place to place causing intermittent disconnections, thus mobility management is required to avoid the discontinuity in services in m-commerce environment. *Geographic location*: the geographic location of a user must be traced to provide suitable location aware services. *Device identity*: mobile devices must be identified in order to cater proper services based on computing and screen capabilities of a device. *Personal identity*: authentication of a user in m-commerce is very much essential, biometric technology may be used for this purpose.

Security: it is a key enabling factor in m-commerce. Compared with the wired counterpart, wireless communications are more vulnerable. The security challenges relate to the mobile device (i.e., should be protected from unauthorized use), the radio interface (i.e., access to telecommunication network requires the protection of transmitted data in terms of confidentiality, integrity, and authenticity), the network operator infrastructure (i.e., security mechanisms for the end users often terminate in the access network) and the kind of m-commerce application (i.e., m-commerce applications, especially those involving payment, need to be secured assures customers, merchants, and network operators).

Business: a key challenge that companies face as they build businesses for the wireless and wired age is that they will need to integrate capabilities and disciplines that are quite separate in most organizations today. These include creative thinking, business skills, a deep understanding of technology and technical issues in both telecommunications and information systems, an understanding of how all this will evolve, and better skills in design and branding.

Negotiation: negotiation in mobile commerce environment is required to reach towards proper agreement while purchasing goods and services using heterogeneous devices and networks. Hence, negotiation is fundamental to the provision and management of m-commerce services in a marketplace model. However, automating this negotiation is a challenge because of the components like agent languages, protocols, etc. A detailed view of negotiation in m-commerce is given next.

NEGOTIATION IN MOBILE COMMERCE

Negotiation in m-commerce environments needs to consider the constraints and requirements imposed by the domain, uncertainty about the opponent's negotiation parameters and inter-dependence between issues. The negotiation is one of the market transaction phases in m-commerce environment. It is the phase in which the sellers and the buyers try to make an agreement over the goods and QoS. In addition, the method of payment and shipping has to be fixed in the agreement. The multi-party negotiation should support the concurrent execution of the services on each server by respecting the consistency of the global execution work (Michaela, 2002; Ying, 2003). Three broad topics for research on negotiation that serve to organize the issues under consideration are identified (Jennings, 1999). They are as follows.

- **Negotiation protocols** are the set of rules that govern the interaction. The permissible types of participants (e.g., the negotiators and relevant third parties), the negotiation states (e.g., accepting bids, negotiation closed), the events that cause state transitions (e.g., no more bidders, bid accepted), and the valid actions of the participants in particular states (e.g., which can be sent by whom, to whom and at when).
- **Negotiation objects** are the range of issues over which agreement must be reached. These may be single issue, such as price, or multiple issues relating to price, quality, timing, etc. Also relevant here are the allowable operations on these objects. In the simplest case, the structure and contents of the agreement are fixed, and negotiation amounts to accepting or rejecting the offer. The next level, however, offers flexibility to change the values of the issues in the negotiation object, through counter-proposals, changing the structure of the negotiation object (by adding guarantees), and so on.
- Finally, the **reasoning models** provide the decision-making apparatus by which participants attempt to achieve their objectives. The sophistication of the model is determined by the protocol used, the nature of the negotiation object, and the range of operations that can be performed on it.

Like all other environments, negotiation in m-commerce environments needs to consider the constraints and requirements imposed by the domain (Andreou, 2005). For example, users on mobile devices would normally be involved in short negotiations whereas those on fixed hosts can take part in continuous and computationally expensive negotiations. Mobile phones require well targeted and concise content presented to the user and therefore the mobile commerce service feature negotiation would not include presenting animations, banners and long lists of results. Moreover, the features of mobile telecommunications determine the suitable negotiation mechanism and are often inter-related; the quality of the network (QoN) may itself be partially defined in terms of the quality of service measurements (QoS). The QoN in turn vary with changes in bandwidth, range, and frequency of disconnections, costs of connection, data integrity and security. Negotiation mechanisms can be designed to adapt to such variations in bandwidth and QoN.

1. The QoN influences rates of concession or its decisions. For example, if the QoN is low then first acceptable offer may be considered, while if the QoN is high, then bargaining could be done to search for the best deal and maximize profit.
2. Bandwidth limitations and fluctuations restrict the number of users involved in a negotiation, the number of messages or the number of concurrent negotiation threads.

3. If a mobile phones or laptops battery power is low, user may decide to concede and quickly find an agreement or send notifications that it will soon suspend or abandon the negotiation.
4. If there is not enough computational memory and processing power, then user can choose not to adopt complex strategies. In case of increased latency and loss of network performance, users may choose to timeout or increase its time to compute its strategies and plans while waiting for a message.

ROLE OF AGENTS IN MOBILE COMMERCE

Software Agents

Software agent is an autonomous software entity that can interact with its environment. In other words, they are agents that are implemented using software. This means that they are autonomous and can react with other entities, including humans, machines, and other software agents in various environments and across various platforms (Manvi, 2004). Agents have certain special properties, which make them different from the standard programs such as mandatory properties and orthogonal properties. Mandatory properties of the agents are: autonomy, reactivity, proactivity, temporal continuity, goal oriented and adaptive. The orthogonal properties are: learning, mobility, communicative, and collaborative (Russell, 2001).

Mobile agents are the agents that can physically travel across a network, and perform tasks on machines that provide agent-hosting capability. This allows processes to migrate from computer to computer, for processes to split into multiple instances that execute on different machines, and to return to their point of origin. The motivation for using the mobile agents is to perform dedicated processing inside a network at particular links/clusters on behalf of customers/administrators, which may not implement required algorithm and services. A major incentive for mobile agent based approach is that policies can be implemented dynamically based on resource states (Reinhard, 2005).

An agent platform comprises of agent server, agent execution environment and transport mechanisms (Menelaos, 1999). The platform offers the following services: persistence, agent mobility, agent creation, inter-agent communication and security. There are several good reasons for using agent mediated m-commerce: reduce bandwidth consumption and network loads, allow dynamic deployment of application components to arbitrary network sites, encapsulation of protocol, execute asynchronously and autonomously, adapt dynamically. The characteristics of agents in m-commerce services are: structures for knowledge representation, responsiveness, adaptivity, sociability and locality of interaction and autonomy.

Agent technology is emerging as a new paradigm in the areas of artificial intelligence and computing which facilities the features for future mobile services and applications development (Griss, 2001). Agents are said to become the next generation components in software development because of its inherent structure and behavior supporting CBSE (Component Based Software Engineering). In m-commerce environment, agents play role in many ways as follows: Agents as mediators, agents as enabler, agents as facilitators, and agents as intermediaries.

AGENT BASED PRODUCT NEGOTIATION MODELS

Agent based negotiation is the process in which two or more parties/agents bargain for goods or information reciprocally. Negotiation between agents will ensure speed, consistency and freedom from human error. Various negotiation models for m-commerce are as follows: Auction Model (like English Auction, Dutch auction, sealed bid auction, etc), trade off based model, argumentation based model, **c**ontract net protocol model, bilateral negotiation and game theory based model etc. Uses of agent based negotiation models in m-commerce environment offers several benefits: cost effectiveness, technology power, value of the Internet access, market share, cellular networks, standardization, and pricing models. In m-commerce environment, agent based negotiation models are used on the basis of device constraints (such as display, memory, speed and so on), network resource constraints (i.e., bandwidth) and mobility constraints.

Agent based negotiation models in mobile commerce differ from e-commerce in the following ways:

- Software agents can consider searching of required commerce services based on wireless network conditions and customer locations,
- Path to the service provider (especially for multimedia applications) can be reserved with resources by agents sent from the customer depending on device and application constraints,
- Software agents can perform trading based on quality of service of the network by varying the incentives offered to the customer.

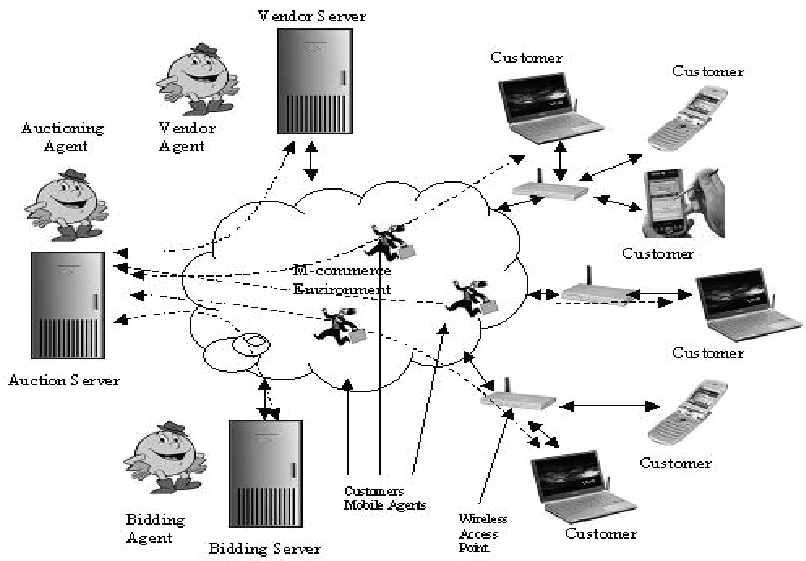
Software agents based negotiation models have the capability of solving number of issues like bandwidth, connectivity, low power processing, resource constraints etc. Agents work with small, specialized tasks and are able to coordinate their work depending on their interpretation of the m-commerce environment. Software agents will enhance the functionality of m-commerce applications and facilitate the user's effort.

In order to utilize the available resources more efficiently, a mobile agent can be used to perform resource allocation to obtain the required QoS for m-commerce services such as product brokering. A mobile agent can suspend the work and move it to a platform in a fixed network (migration), thus outsourcing resource-expensive work, and eliminating the need for a persistent connection. Network limitations and disconnections in m-commerce are solved by the autonomous work, the agent returns with a result when the connection is restored. Existing mobile processors are capable of displaying video at low data rate, but not good enough for most m-commerce applications. Software agent based approaches may be a solution for low power requirements to have them running 24x7 with at least 24 hours battery life.

The success of agent based product negotiation models depends on basic telecommunication facilities (QoN) provided like bandwidth, network connectivity, device constraints etc. In each agent based negotiation models the QoN (i.e. bandwidth) influences rates of concession or its decisions in negotiation of products between the agents in each model. For example, if the QoN is low then first acceptable offer may be considered, while if the QoN is high, then bargaining or negotiation could be done to search for the best deal and maximize profit in each agent based negotiation models.

In each model, bandwidth limitations and fluctuations restrict the number of users involved in a negotiation. If a mobile phones or laptops battery power and bandwidth is low, user may decide to concede and quickly find an agreement or send notifications that it will soon suspend or abandon the negotiation. Even the problem of negotiation arises not only among sellers and bidders about goods to exchange but also among different kind of users, who try to access network services and negotiate to fairly share bandwidth. Therefore, in case of increased latency and loss of network performance, incompatible resolution and etc users may choose to timeout or increase its time to compute its strategies and plans while waiting for a message.

Figure 1. Auction based negotiation model



AUCTION BASED NEGOTIATION MODEL

The auction model supports a negotiation mechanism for selling goods to bidders. Normally, an auction is a one-to-many negotiation, between one seller agent and many buyer agents. But in some cases, such as Double auction, seller agents and bidder agents could be M: N. The auction type's that are most widely used in online auctions and analyze them from the point of view of agents negotiation is presented in (Timmers, 1998; Beam, 1997).

Figure 1 illustrates the agent's usage in auction negotiation model. It comprises of bidding server, auction server, vendor server and customers. We can see in the figure that mobile agents are sent from some of the customers to auction servers to attend the auction and bid on behalf of the users in m-commerce environment. Bidding agents (static agents) bid for products by communicating with auctioning agents. Auctioning agents are employed to auction the products, and the mobile agent sent from the customer will participate in the auction as per the intelligence programmed by the agent creators. Sometimes, the auctioning agent can contact its creator, if it wants to take its own decision on the auction price. Similarly mobile agents can participate in bidding and negotiating for the auctioning products i.e. price negotiations or services through mobile devices. The negotiation process exists between auction agents and vendor agents as well as bidding agents. The vendor agent negotiates for the auctioning products and services.

Negotiation is considered in all types of auctions that include English, Dutch, Sealed bid auction and Vickrey auction. In the English auction, the buyer agents successively raise the bids up to their reservation prices until only one-buyer remains. In a standard terminology, the buyer agent reservation price is the maximum price he is willing to pay for the item and the seller agents reservation price is the lowest acceptable sale price for the item. Then, if a buyer agents maximum bid meets or exceeds the seller agents reservation price, the transaction between the seller agent and the buyer agent is completed. Sometimes, more than one buyer agents price is same in auction, and then negotiation process exists between the buyer agents and seller agents. The rules of the English auction are easy to understand and to implement in the framework of an online auction. As a result, this auction type is the most popular among online auctions both without and with agent support.

In the descending bid auction (Dutch auction) the seller agent starts at a very high price, and then the price is decreased gradually until a buyer agent accepts the asking price. All buyer agents can see the current price and then decide if the price is still too high or if they wish to purchase at that price. The first bidder agent at the current price wins the auction.

In the first price sealed bid auction, each buyer agent independently submits a secret single bid. Later, bids are opened simultaneously and the item is sold to the buyer agent who makes the highest bid. Nobody is allowed to update a bid once submitted, and the winner pays the highest price bid.

The Vickrey auction is similar to the first sealed bid auction. Each buyer agent independently submits a single bid, without seeing bids of other buyer agents, and the winner is the buyer agent with the highest bid. However, the price paid by the winner is the second-highest bid.

Auction theory provides us with an understanding of the conditions under which the various auction types are optimal and which strategies are optimal for participants in different auctions. The work given in (Ibrahim, 2004) describes about the agent-based auctions through mobile devices. It describes the process of auctioning goods and services using mobile devices in a flea market like scenario. It elaborates on a prototypical process how these auction methods; interaction requirements and other attributes can be incorporated in an agent based mobile auction.

The work in Kasbah (Chavez, 1997; Chavez, 1996) is a web based and agent based consumer-to-consumer electronic marketplace. Users intending to buy or sell goods can create buying and selling agents that help them transact their product. When creating agents, users must give them some strategic direction, and afterward send them off into a centralized agent marketplace. Kasbah agents proactively seek out potential buyers or sellers so as to negotiate with them on their owners' behalf. Thus these agents are responsible for automating much of the merchant brokering and negation stages for both buyers and sellers. Each agent arriving into the central marketplace is targeted to close acceptable deal according to a set of user-specified constrictions such as desired price, lowest (or highest) acceptable price and a deadline to complete the transaction.

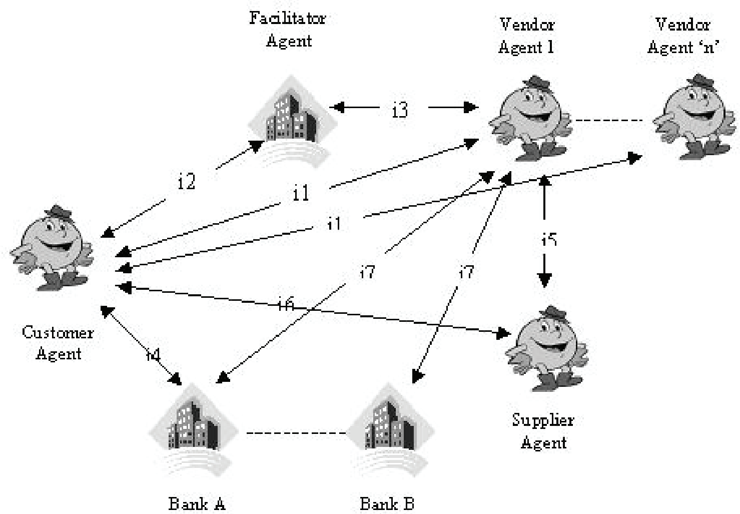
The work given in (Benchaphon, 2000) presents a brief overview of the electronic commerce transaction process and particularly the negotiation process. Agent based negotiation model includes the capability of bargaining on other product offers rather than only on the prices. The work given in (Mathieu, 2002) presents a generic model for contract negotiation. It defines a generic agent state model and a generic protocol for contract negotiations, called ANTS and proposes a three level methodology to build generic negotiation applications based on a communication protocol, a generic internal structure for problem representation and a strategy level to obtain better rationality. It is possible to describe precisely a generic model that we can use in several negotiation and marketplace problems.

Trade off Negotiation Model

Trade-offs is fundamental to decision-making problems in business, especially to dynamic decision making problems and negotiation problems. Negotiation criteria, duration and complexity vary depending on the nature of business and the type of transaction. Further, negotiation often may need to be based on several parameters/attributes rather than price alone (e.g. quality, delivery conditions, and warranty, after sales service, loan options and other value added services).

Figure 2 shows the trading process of negotiations in various situations of participants or agents like – customers, vendors, suppliers, banks and

Figure 2. Trade off negotiation model



facilitators in trade off negotiation. The arrows represent information flows between entities as they negotiate with each other for services. Number denotes the interaction sequence.

The sequence of interactions is described as follows:

1. Buyer or customer agent is on a train and his agent, acting as a customer on a mobile device, browses and negotiates the offerings of one or more vendor agents for purchasing flight tickets or products (interaction i1).
2. Customer agent, (from his mobile phone on a train) negotiates for the brokerage services of a facilitator (interaction i2*)* by deliberating on the facilitators cost, effort to be spent on finding vendors and the deadline for finding them. Customer agent provides the transaction requirements to the facilitator agent, such as departure point, time, destination and maximum price.
3. Facilitator agent negotiates for the services of various vendor agents for information about flights for the requested route (interaction i3*)*. The facilitator agent selects vendor agent1 as offering the most suitable package and further negotiates with it. Facilitator agent submits customer agent payment credentials to vendor agent1.
4. Customer agent negotiates for a loan with a banker at financial bank 'A' (interaction i4*)* regarding the amount, interest rates and time of repayment. The banker is in a seminar and his agent is active on his laptop via a wireless link. Since customer agent has a good bank record and the loan is not high, the banker's agent does not alert its owner.
5. Vendor agent1 negotiates with the supplier agent on how to deliver the tickets or products (interaction i5*)*. The supplier agent negotiates with customer agent over the most appropriate way and time to deliver the tickets or products (interaction i6*)*.
6. Vendor agent1 compensates facilitator agent for the completed transaction (interaction i3*)*.
7. Vendor agents negotiate with financial banks for customer services (interaction i7*)*.

As can be seen, the interaction between the participants in this scenario incorporates negotiation at various stages. Some of the participants are using mobile devices and are located on trains, walking or in a seminar. To support such a scenario we require dynamic trading of m-services. To achieve such personalization, negotiation can be used in the trading of both telecommunications services and high-level services between participants or agents in an open electronic market.

The work given in (Xun, 2000) illustrates the secure agent-based framework for Internet trading in mobile computing environments. Most of the current Internet trading frameworks, in particular their negotiation and payment phases, are intended for customers frequently connected to the Internet during an entire transaction. This requirement cannot be easily met in the high communication cost and/or low bandwidth settings, typically found in mobile computing environments. The framework is composed of two new protocols. One is the agent-based auction like negotiation protocol; another is the agent-based payment protocol. Both of them are dedicated to solve the trade problems of Internet trading in mobile computing environments and ensured to be safe by cryptographic technologies. The combination of the two secure protocols constitutes an integrative solution for Internet trading in mobile computing environments.

Argumentation Based Negotiation Model (ABN)

Argumentation is a type of negotiation that allows agents to exchange meta-information such as justifications, critics, and other forms of agent interactions. Agents can increase the likelihood and quality of an agreement by exchanging arguments, which influence each other's states. The basic negotiating agent should provide a generic description of agent's capability of conducting argumentation-based negotiation. An argumentative negotiator shares many components with the basic negotiator. For example, it needs to be able to evaluate proposals, generate proposals and so on. What makes argumentative agents different is that they can exchange meta-information or arguments in addition to the simple proposal, acceptance, and rejection utterances. These arguments can potentially allow agent to (i) justify their negotiation position; or (ii) influence the counterparty's negotiation position. This may lead to a better chance of reaching agreement and/or higher-quality agreements.

In ABN, influencing the counterparty's negotiation position takes place as a result of providing it with new information, which may influence its mental attitudes (e.g., its beliefs, desires, intentions, goals, and so on). This might entice (or force) the agent to accept a particular proposal, or concede on a difficult issue. Arguments can range from threats and promises to logical discussion of the agent's beliefs. In order to be capable of engaging in ABN, an agent needs the following additional capabilities:

1. **Argument Evaluation:** this component encompasses the ability of the agent to assess an argument presented by another, which may cause updates to its mental state. This is the fundamental component that allows negotiators' positions to change.
2. **Argument Generation:** this component allows the agent to generate possible arguments, either to support a proposal, or as an individual piece of meta-information.
3. **Argument Selection:** Sometimes, there might be a number of possible arguments to present. For example, an agent might be able to either make a promise or a threat to its opponent. A separate component is needed to allow the agent to choose the more preferred argument. Selection might be based on some analysis of the expected influence of the argument, or on the commitments it ties the utterer to.

The work given in (Sierra, 1998) illustrates about the autonomous agents usually operate as a multi-agent community performing actions within a shared social context to achieve their individual and collective objectives. Argumentation-Based Negotiation (ABN) has been advocated as a promising means of resolving conflicts within such agent societies. The work given in (Jennings, 1998) describes the general framework for negotiation in which agents exchange proposals backed by arguments, which summarize the reasons why the proposals should be accepted. The argumentation is persuasive because the exchanges are able to alter the mental state of the agents involved.

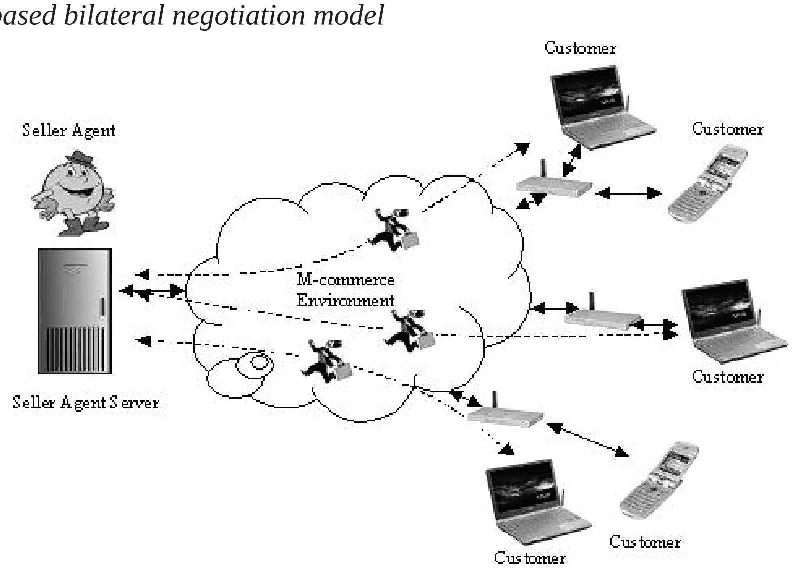
The work given in (Collins, 1998) presents a framework for argumentation-based negotiation. Many autonomous agents operate in domains in which the cooperation of their fellow agents cannot be guaranteed. In such domains, negotiation is essential to persuade others of the value of cooperation. Agents exchange proposals backed by arguments that summarize the reasons why the proposals should be accepted.

Contract Net Protocol/Tendering Model

The contract net protocol is used to allow the agents to a coordination method for task allocation, providing dynamic allocation and natural load balancing. The approach is quite simple and can be efficient. The Contract Net Protocol was first proposed by Smith (Smith, 1980) and demonstrated on a distributed sensing system. The contract net model is described as negotiation though it is simply a very straightforward protocol: task announcement, make bid, award contract etc. For a negotiation to be completed successfully all parties (i.e. agents) must clearly understand the rules of engagement or negotiation protocol. In a simple contract-net protocol, in which a manager agent issues a call for proposals and waits for a full set of replies (or timeouts), each bidder agent must be prepared to honour its bid for the duration of the bid's validity. Otherwise, acceptance of the bid will require a second negotiation, which may itself succeed or fail. If an agent is no longer able to honour its bid, the other agents that originally bid may well have already have accepted other commitments and no longer are in a position to accept that commitment. The whole bidding process might then have to be restarted.

The work given in (Faratin, 1999) presents negotiation architecture for agents that is related to the contract net approach, but extends this model by allowing more elaborate offer/counter interaction sequences between agents. The work given in (Chun, 2003) presents an adaptive

Figure 3. Agent based bilateral negotiation model



negotiation framework for agent based dynamic manufacturing scheduling. Issues regarding adaptive negotiation are addressed at both the system architectural level and the agent model level. Different economic models and their characteristics are compared and selection heuristics are discussed as well.

Bilateral Negotiation Protocol Model

Bilateral negotiation is a process of making offers and counter offers with the aim of finding an acceptable agreement.

Figure 3 depicts the agent based bilateral negotiation protocol model in mobile commerce environment. The bilateral negotiation process is straightforward. For example, a customer wants to buy a TV set. He /she logs into the m-marketplace and inputs his/ her requirements for the TV. Mobile agents are sent from some of the customers to seller agent server to attend the negotiation for the products on behalf of the users to exchanging the offer and counter offers. In which, buyer agent or customer may receive offers/proposals from sales agents. Each proposal defines, for example, a complete product offering including the size of the TV set, brand, price, the quality of the TV and return policy etc. The buyer agent evaluates these proposals based on its buyer's criteria and makes a suggestion (counter offer). The buyer's agent continues negotiation until an agreement can be reached. On the other hand, if the buyer and seller agents reject the current incoming offer criteria and conditions, the negotiation terminates at the 'no deal' state.

The work given in (Raz, 2007) outlines an automated agent design for bilateral negotiation with bounded rational agents where there is incomplete information regarding the opponent's utility preferences. The automated agent incorporated a qualitative decision making mechanism. The results showed that agent is indeed capable of negotiating successfully with human counterparts and reaching efficient agreements.

Game Theory Based Negotiation Model

Game theory-based negotiation involves the application of concepts such as utility functions, space of deals and strategies and negotiation protocols. Agents use payoff matrices to represent common knowledge (each agent knows the utility value of the outcome of some interaction). The game theory is used to achieve coordination among autonomous agents in cooperative domains.

Game theory is a mathematical framework designed for analyzing the interaction between several agents whose decisions affect each other. The game theory is widely acknowledged to provide a useful set of tools for the design of multi-agent architectures (Binmore, 1999). Framework of game theory consists of two main classifications: cooperative and non-cooperative game theory. A game is cooperative if the players or participating agents are able to form binding commitments. Non-cooperative game theory deals largely with how intelligent individuals interact with one another in an effort to achieve their own goals. Game theory has been for a long time the subject of study by researchers interested in competitive decision-making, bargaining and negotiation.

Game theory models have provided a useful framework for the coordination of rational agents in conflict resolution. Two fundamental concepts have presided over the application of game theory to automated negotiation - Nash Equilibrium and Pareto efficiency. Nash equilibrium is obtained when two agents have no longer any incentive in deviating from their strategies. The convergence afforded by Nash bargaining game equilibrium makes it the most popular solution to the bargaining problem in contrast to the Bayesian equilibrium concept, which was adapted to games with incomplete information (Gerding, 2000).

One way to describe a game is by listing the players (or individuals) participating in the game, and, for each player, listing the alternative choices (called actions or strategies) available to that player. In the case of a two-player game, the actions of the first player form the rows and the actions of the second player the columns of a matrix. The entries in the matrix are two numbers representing the utility or payoff to the first and second player, respectively. A play consists of choosing certain strategies by the players; an outcome of the play is a pair of numbers representing the utilities of the players. To determine "rational" outcomes non-cooperative game theory defines the notion of an equilibrium strategy. Among most widely used concepts of equilibrium strategies are the Nash equilibrium and 'dominant' strategies. A dominant strategy is optimal for all players independent of what the strategies of the other players are. The strategies of players are in Nash equilibrium if no player can benefit by unilaterally changing his strategy.

One of the limitation appearing by applications of game theory is that, at least in classical game theory, frequently simplifying assumptions are made for analysis of a game like the assumption about the full rationality of the players and the assumption that the rules of the play and the preferences and beliefs of the players are their 'common knowledge'. Such assumptions limit the practical applicability of game-theoretic results. In particular, private information such as reservation prices, preferences for different features of the products and the relative importance of these futures as well as time preferences and limitations are usually hidden from the opponent in real-life negotiations.

In the past decades a variety of negotiation models have been developed in the framework of game theory. In particular, a lot of research was made in the fields of cooperative and non-cooperative bargaining with incomplete and vague information, and bargaining over multiple issues.

FUTURE TRENDS

Mobile business strategies explore the new mobile world, look into the future and consider the emerging trends. It discusses the roles of negotiation models between agents, operators, and other key parties in the mobile commerce value chain. Core technologies are addressed from a strategic perspective, familiarizing the reader with both the possibilities and the limitations of the mobile environment. Future negotiation models shall be more intelligent and adaptive in a real world m-marketplace. Cognitive agent architectures that have human kind of reasoning along with game theory based mechanisms could be considered for negotiation in a virtual market that perform bilateral trading, auctioning, etc., in even high mobility conditions and vulnerable wireless environment areas. Several issues arise in such a case: how to build cognition in an agent?; how to address the agent databases and perform communication?; how to provide interaction among agent societies that have information about the market place?, etc.

SUMMARY

M-Commerce is an evolving area of e-commerce, where users can interact with the service providers through a mobile and wireless network, using mobile devices for information retrieval and transaction processing. This chapter discussed the various issues in m-commerce and in particular about the product negotiation models based on software agents. The models discussed are: auctions, trade-offs, argumentation, bilateral, contract-net and game theory. The individual character of future services demands highly flexible service infrastructures and development frameworks. Agent technology can offer a new paradigm for communication over heterogonous network channels and facilitate m-commerce services. Agents provide more flexible and adaptable m-commerce services and also support component based software engineering features such as software reuse, maintainability, customization, and encapsulation**.**

REFERENCES

Andreou, A. S. C., Chrysostomou, C., Leonidou, S., Mavromoustakos, A., Pitsillides, G., Samaras, C., Schizas (2002). Mobile commerce applications and services: A design and development approach. *The First International Conference on Mobile Business M-Business 2002*, July 8-9, Athens, Greece.

Beam, C., & Segev, A. (1997). Automated negotiations: A survey of the state of the art. *Wirtschafts - informatik, 39(3)*, 263-268.

Beer, B., d'Inverno, M., Luck, M., Jennings, N., Preist, K., & Schroeder, M. (1999). Negotiation in Multi-Agent Systems. *The Knowledge Engineering Review***,** *14(7)*, 285-289.

Benchaphon, L., Zhang, Y., & Zhang, Z. (2000). An Agent-based Negotiation Model supporting transactions in Electronic Commerce. *Proceedings of the 11th International Workshop on Database and Expert Systems Applications* (pp. 440 – 444). IEEE Computer Society, Washington, DC, USA.

Binmore, K., & Vulkan, N. (1999). Applying game theory to automated negotiation. *Netnomics, 1(1)*, 1-9.

Chavez, P. M. (1996). Kasbah: An Agent Marketplace for Buying and Selling Goods. *Proceedings of the 1st International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM'96)* (pp. 75-90). Practical Application Company, London, UK.

Chavez., A., Dreilinger, D, Guttman, R., & Maes, P. (1997). A Real Life Experiment in creating an Agent Marketplace. *Proceedings of the 2nd International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, 1198*, 160-179. UK: Springer-Verlag London.

Chun, W., Weiming, S., & Hamada, G., (2003). An adaptive negotiation framework for agent based dynamic manufacturing scheduling. *Proceedings of IEEE SMC 2003* (pp. 1211-1216). Washington DC.

Faratin, P., Sierra, C., Jennings, N., Buckle, P. (1999). Designing Responsive and Deliberative Automated Negotiators. *Proceedings of the AAAI Workshop on Negotiation: Settling Conflicts and Identifying Opportunities* (pp 12–18). Carles Sierra's publications. Orlando, Florida, USA.

George, S. (2002). Mobile Commerce: Vision and Challenges. *Proceedings of the Symposium on Applications and the Internet (SAINT.02.)* (pp. 43-45). IEEE Computer Society, Washington, DC, USA.

Gerding, E. H., van Bragt, D. D. B., & La Poutre, J. A. (2000). Scientific approaches and for negotiation. A game theoretic and artificial intelligence perspective. *CWI report on negotiation*, http://www.cwi.nl/projects/TA/.

Griss, M. L., & Pour, G. (2001). Accelerating development with agent components. *IEEE Computer mag.*, *34(5)*, 37-43.

Irvine, C. (2001). Emerging Value Propositions for M-Commerce. *Journal of Business Strategies, 18(2)*, 133-149.

Ishara, K. S. (2003). *Mobile Commerce- Usage and Usability*. Retrieved from http://General.rau.ac.za/infosci/www2003/Papers/Singh,%20IK%20Mobile%20commerce%20-%20usage%20and%20usability.pdf

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of ACM*, *44(4)*, 35-41.

Jennings, N. R., Parsons, S., Noriega, P., & Sierra, C. (1998). On argumentation Based Negotiation. *Proceedings of the Int. Workshop on Multi-Agent Systems*. (pp. 1-7). MIT.

Keng, S., & Ee-Peng, L. (2001). Mobile Commerce: Promises, Challenges, and Research Agenda. *Journal of Database Management, 12(3)*, 4-24.

Khalil, I., Ibrahim, R., & Kronsteiner, G. K. (2004). Agent-Based Mobile Auctions: The Flea Market Scenario. *Proceedings of 2004 International Research Conference on Innovations in Information Technology (IIT2004),* Dubai.

Kraus, S. (1992). Negotiation in a non-cooperative environment. *Journal of Experimental and Theoretical Artificial Intelligence*, 3(4), 255-281.

Kronsteiner, R., Khalil, I., & Ibrahim, G. K. (2005). Agents for Mobility Engineering in m-Business. *ACM International Conference Proceeding Series; Proceedings of the 4th international symposium on Information and communication technologies, session Information technology, 92,* 57-62. Trinity College Dublin.

Manvi, S. S., & Venkataram, P. (2004). Applications of Agent Technology in Communications: A Review. *Computer Communications, 27(15)*, 1493-1508.

Mathieu, P., & Verrons, M. H. (2002). A generic model for contract negotiation. *Proceedings of the Artificial Intelligence and the Simulation of Behaviour symposium on Intelligent Agents in Virtual Markets (AISB'02), 3*(3), 1-8.

Menelaos, K. P., Fotis, G. C., Iakovos, S. V., & Gennaro, M. (1999). Mobile agent standards and available platforms. *Computer networks, 31(19)*, 1999-2016.

Michaela, O. (2002). An Adaptive Negotiation Model for Agent Based Electronics Commerce. *Studies in Informatics and Control, 11(3)*, 45-46.

Raz, L., Sarit, K., Wilkenfeld, J., & Barry, J. (2007). An *Automated Agent for Bilateral Negotiation with Bounded Rational Agents with Incomplete Information*.

Rosenschein, Z. (1989). Negotiation and task sharing among autonomous agents in cooperative domains. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 912-917), Morgan Kaufmann, San Mateo, CA.

Russell, S., & Norvig, P. (2001). Artificial Intelligence a modern approach. Prentice Hall series in AI, India.

Schwiderski, S., Heiko, G. K. (2002). *Secure M-Commerce*.

Shaffer, R. A. (2000). M-Commerce: Online Selling's Wireless Future.

Sierra, C. N, Jennings, R., Noriega, P., Parsons, S. (1998). A framework for argumentation-based negotiation. *Proceedings of ATAL 1997, 1365,* 177-192. Berlin, Germany: Springer-Verlag.

Smith, R. G. (1980). The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computer, 29*, 1104-1113.

Tarasewich, P., Nickerson, R. C., & Warkentin, M. (2002). Issues in Mobile Commerce. *Communications of the Association for Information Systems*, *8,* 41-64.

Timmers, P. (1998). Business Models for Electronic Markets. *Electronic Markets, 8*(2), 3-8. www.it.iitb.ac.in/~annanda/papers-280905/Secure%20m-commerce%20ECEJ.pdf

Xun, Y., Chee, K., Siew, X., Feng, W., & Eiji, O. (2000). A Secure Agent-based Framework for Internet Trading in Mobile Computing Environments. *Distributed and Parallel Databases, 8,* 85-117. MA, USA: Kluwer Academic Publishers Hingham.

Ying, H., Ching, W., Wang, L., Hua, H., Wen, H., & Lo, H. (2003). Applying Mobile Agents to E-business. *Tamkang Journal of Science and Engineering*, *6*(3), 159-172.

WEB SITES

http://search.epnet.com/athens.asp

http://www.dcs.kcl.ac.uk/staff/aspassia/AISB.html

http://www.umiacs.umd.edu/~sarit/Articles/linetalEcai06.pdf

KEY TERMS

Argumentation Negotiation: Argumentation is a type of negotiation that allows agents to exchange meta-information such as justifications, critics, and other forms of agent interactions.

Auction: Auction is a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants.

Bilateral Negotiation: Bilateral negotiation is a process of making offers and counter offers by the two parties of agents with the aim of finding an acceptable agreement through mobile devices in mobile commerce environment.

Contract Net Negotiation: The contract net protocol is an automated negotiation for distributed problem solving and electronic marketplace for selling and buying of goods. The contract net protocol is used to allow the agents to a coordination method for task allocation, providing dynamic allocation and natural load balancing.

Game Theory Based Negotiation Model: Game theory is a mathematical framework designed for analyzing the interaction between several agents whose decisions affect each other. The game theory is widely acknowledged to provide a useful set of tools for the design of multi-agent architectures.

Mobile Agents: Mobile agents are the agents that can physically travel across a network, and perform tasks on machines that provide agent-hosting capability.

Mobile Commerce: Mobile commerce (m-commerce) is defined as the use of wireless terminals, such as a cellular telephone or PDA, and a network to conduct business transactions, exchange of information, buying/selling of services or goods, etc.

Negotiation: The goal of the negotiation process is to achieve an agreement, which is mutually acceptable among parties or agents/users over certain goods or services.

Software Agent: Software agent is an autonomous software entity that can interact with its environment.

Tradeoff Negotiation: Trade-offs are fundamental to decision-making problems in business, especially to dynamic decision making problems and negotiation problems.

Chapter LIII
Role of Telecommunications in Precision Agriculture

James M. McKinion
USDA-ARS, USA

ABSTRACT

Precision agriculture has been made possible by the confluence of several technologies: geographic positioning systems, geographic information systems, image analysis software, low-cost microcomputer-based variable rate controller/recorders, and precision tractor guidance systems. While these technologies have made precision agriculture possible, there are still major obstacles which must be overcome to make this new technology accepted and usable. Most growers will not do image processing and development of prescription maps themselves but will rely upon commercial sources. There still remains the challenge of storage and retrieval of multi-megabytes of data files for each field, and this problem will only continue to grow year by year. This chapter will discuss the various wireless technologies which are currently being used on three proof-of-concept farms or areas in Mississippi, the various data/information intensive precision agriculture applications which use wireless local area networking and Internet access, and the next generation technologies which can immensely propel precision agriculture to widespread use in all of agriculture.

INTRODUCTION

Multispectral image-based precision agriculture technology is beginning to have widespread use in row crop production agriculture in the United States, particularly in the cotton belt. Companies such as InTime, Inc.[1] (InTime, 2007), located in Cleveland, Mississippi, USA, are providing image-based products from which clients have access to scout maps derived from multi-spectral images. InTime uses Geospatial System's Inc. multispectral image cameras mounted on fixed-winged aircraft to obtain their image information (GSI, 2007). Specific scout maps target different plant or soil characteristics such as overall plant biomass, differences in soil type, and differences in fertilizer nitrogen uptake. The scout maps are used to determine rates of plant growth regulators, insecticides, defoliants, herbicides, or fertilizer to apply to the plants. Utilizing InTime's Web-based Crop-Site, growers and consultants can easily transform their scout maps into vehicle/controller prescription files that allow for chemical rates to be varied automatically with minimal operator inputs.

The information products generated by such activity, as described previously, can easily be expressed as multi-megabyte sized data files, especially when geographic information system (GIS) technology is used. In almost all situations, multispectral-image based maps are geo-referenced with pixel sizes in the 0.5 meter to 1 meter range generating large data files for the applications derived from these maps. The application maps are also generally geo-referenced for use by the geographic positioning systems (GPS)-based controllers on the farm equipment. For a 500 ha field, application maps can easily be generated which are from 1 to 5 megabytes in size or larger.

Many controller manufacturers today use PC cards or similar technology which has to be hand-carried from the farm base of operations to the equipment in the field and inserted into the controller (Raven Industries, 2008). Conversely, after the application has been made by the equipment (planting, fertilizer application, pesticide application, etc), the PC cards have to be manually collected and taken to the operations base to have the as-applied map uploaded into the farm computer. While many medium and small farms are not affected by this information movement process since the farm manager/owner is also the equipment driver, larger farms which have many pieces of equipment are often scattered over 50 km from one side of the farm land to the other, or even greater distance. The distances involved from the farm base station to the equipment in the field presents an operations problem for growers since when they are involved with precision applications not only must they travel to and from the fields where the equipment is located they must also find the equipment in the field which can often be two or more kilometers across to deliver and pick up the PC cards containing information. This type of operation can easily take a person three or more hours per day just to deliver and/or pick up the data cards.

When time critical operations are involved in delivery of pesticide application maps to the spray equipment controller, this problem becomes even more exacerbated. Our research in early-season plant bug control has shown that from the time the multispectral image is taken by the airplane to the time the spatially variable insecticide is delivered to the spray equipment no more than 48 hours should elapse (Willers et al., 1999). The optimal time is no more than 24 hours. Obviously, there is a better way to solve this time constraint/labor problem than using people for hand carrying PC cards.

Wireless local area network technology is a practical solution to movement of information to and from farm machinery which use GPS-guided precision application controllers. Low-cost wireless network solutions are available and are beginning to see widespread use in the United States. Commercial cell telephone networks are beginning to offer medium–speed Internet access via their cellular telephone towers.

Another technology which is becoming very widely used in the United States is the use of precision guidance on farm application machinery. This technology has proved it worth in labor savings alone by removing the tedium from equipment operators so that they can monitor the application operation to ensure seeds or chemicals are being applied as the equipment moves across the field, allowing the operator to stay longer in the field without the exhaustion which accompanied pre-precision guidance operations. Another tangible benefit from precision guidance cultivation is the recovery of lost row acreage which has been estimated as high as 10% of large field area. This recovery for cotton production on larger plantations has been estimated to pay for the precision guidance system in as little as one year. Many precision guidance systems use the 900 MHz public spectrum to send the signals from the GPS base station to the farm equipment to allow up to centimeter accuracy in the placement of farm application machinery in operation.

This chapter will discuss the various wireless technologies which are currently being used on three proof-of-concept farms or areas in Mississippi, the various data/information intensive precision agriculture applications which use wireless local area networking and Internet access, and the next generation technologies which can immensely propel precision agriculture to wide spread use in all of agriculture.

CURRENT STATE OF AFFAIRS

All farms engaged in row crop production agriculture have extensive information available on their farm operations, crop inputs in terms of seed used and fertilizers applied, soil types, soil fertility, pesticides applied and weather information. Some farms are more organized than others and have all this information developed in computerized databases. Modern production agriculture requires that the aforementioned information and much more be available for not only pre-season and in-season decision making, but also for many regulatory purposes required by state and federal government agencies and by seed companies which provide genetically modified varieties used in current practices, especially in cotton production.

The National Cooperative Soil Survey began more than a century ago. Soil databases began to be accumulated by the US Department of Agriculture for the entire US in earnest in the 1940's, 50's and 60's in terms of soil maps whereby soil types were identified and classified by agents of the Soil Conservation Service. These agents would walk the fields, take soil cores manually, and identify broad areas on soil maps by soil type. Today, one can obtain detailed soil maps for each county in a state and generally rely on the placement of these soil types within crop fields. Since August 2005, soil maps and associated data and information have been available through a geographic information system called the Web Soil Survey (WSS). The newest version of the WSS is at http://Websoilsurvey.nrcs.usda.gov.

In the past, soil fertility has been obtained by taking representative soil cores from each major soil type within a field, where soil type was obtained from the maps mentioned previously, and sending the soil samples to a state or private laboratory for analysis to determine N, P and K content.

Yield histories on a field or worst case farm basis have been obtained for management purposes. The advent of private and commercial access to the Global Positioning System (GPS) has changed not only the type of record keeping but has made the level of record keeping much more detailed. Most of the data records were paper records of field histories, planting data, fertilizers applications, chemical applications, fuel usage, equipment usage, and yield data which were kept on a field or farm basis. Only in the last 15 years have these data been migrated from the paper environment to the computer database environment.

Early use of telecommunications on farms was primarily limited to farm headquarters communicating to equipment operators and farm personnel using citizens-band radios, licensed two way radios, and paging equipment.

GPS AND PRECISION STEERING

In the 1970s and 80's the US Military began launching a series of 24 satellites into low orbit above the earth with a pattern of orbits such that most of the time at least 4 to 6 satellites are in line-of-sight to an observer on the earth's surface (Daly, 1993). These satellites emit radio signals with extreme precision time information so that with appropriate receiving equipment, one can determine one's location on the surface of the earth. This system is known as the Global Positioning System or GPS. Today using what is known as differential signal correction one can determine one's position within plus or minus one meter of the true location with hand-held devices. Going even further, when a secondary radio system is used with GPS signals and the secondary system's location is precisely known, position accuracies within plus or minus one centimeter can be determined. This later system is known as a survey-grade GPS system and is the basis for auto-steering technology which is used in precision guidance of farm machinery.

In the 1970's and 80', software began to be developed to take advantage of satellite imagery so that precision maps could be obtained and manipulated. As GPS became available, this feature was integrated into the software systems so that ground features collected manually using GPS sensors could be added to the maps. This allowed the soil maps which were collected by the Soil Conservation Service and the US Geological Survey to be digitized and converted into spatially sensitive computer maps. Several break thoughs happened during this time which affected the use of the data contained in the old soil maps. Land-leveling was found to be advantageous especially concerning the application of irrigation technology (Walker, 1989). Land-leveling required the use of laser technology (one-way communication) to automatically raise or lower machinery used in the leveling process. This process also changed the morphology of the soil types in the field so that field had to be re-sampled to determine their new soil type. GPS technology was used to map the boundaries of the soil types in the fields to much higher level of accuracies. Millions of acres of land were land-leveled. Even farms that were not land-leveled began to have their field re-sampled using GPS to determine more accurate soil-type boundaries within the fields.

Within the last 10 years, precision guidance or auto-steering systems have become widespread in use (Lessiter, 2006). These systems make use of GPS and an on-site beacon system typically using the 900 MHz unlicensed spectrum to broadcast location and timing signals to field machinery used in precision operations. A typical auto-steer system will have two to three satellite GPS receivers mounted on the field vehicle and a receiver antenna for the geo-referenced beacon system, which can be as far as 50 km from the machinery. These systems typically use frequency hopping radios and a fairly low transmission rate such as 125 to 250 kbps. According to the physics of signal propagation, the lower the transmission rate the further the signal propagates. Again this type of system is a one-way transmission from the satellites and the beacon down to the receivers on the field equipment. Signals from the auto-steer system allow tractors, sprayers, combines, and harvest equipment to navigate a predetermined path with centimeter accuracy. Auto-steer systems allow extended hours of operation by operators without the fatigue associated with non-auto-steer vehicles. Operators can pay more attention to the operation of the equipment to assure the operation is being carried out properly, i.e., seed planting equipment or chemical applicators are working properly and not clogged. Another huge benefit of using auto-steer systems is the recovery of lost acreage. Conventional row crop operations lose area because the operator can not precisely steer the equipment 100 % of the time so rows are made with less precision, typically loosing 10% of the field. This 10% can now be recovered and made productive. Large cotton plantations typically can pay for the cost of auto-steer systems within the first year of operations just from the recovery of lost acreage and the increase in productivity from that area.

REMOTE SENSING VIA AIRCRAFT

Another facet of the precision agriculture technology package is the availability of timely images of row crops within the growing season. Because of the timeliness factor, satellite imagery is of very limited to no use at all, because from the time the image is acquired to the time it is made available for decision making months have elapsed. Imagery acquired via fixed-winged aircraft has become commercially available widely in the US over the last five years. Companies such as InTime, Inc can provide 24 to 48 hour response time to image requests making this technology available for rapid response pest management as well as longer term crop management decisions. Images acquired for use in crop and pest management typically involve four layer images using red, blue, green, and infrared filtered images which are spatially and geo-rectified. From these four layers, image analysis software such as Imagine and ARCView can be used to construct prescription maps which can be used to make decisions about fertilization, irrigation, pesticide applications, application of plant growth regulators, and harvest aid chemical applications. Since most farms are by definition remotely located and do not have access to the Internet, these images and application maps currently have to hand-carried to the farm headquarters for distribution to farm managers and crop and pest consultants. The lack of high speed digital data communication is a major bottleneck in the promotion and acceptance of image-driven precision agriculture technology. This scenario, however, is beginning to change.

PRESCRIPTION FARMING

Even without the help of remote sensing and imaging technologies, precision agriculture has been in use for over 15 years and is becoming wide spread in use. The advent of the low cost micro-controllers based on the microcomputer has made this possible. With computer technology becoming available on the farm beginning in the middle 1980's, farm record keeping was a natural advance from paper records to digital database technology. As computers became more and more powerful and hard discs became ever larger in size, the computer became more than just a good way to keep records to satisfy income tax requirements and EPA regulations on chemical and pesticide usage. Software became available in the early 1990's which allowed farm managers the capability to specify planting density, chemical application rates, and fertilizer application rates which corresponded to their historical records of soil type, soil fertility, cation exchange capacity, field yield data, and weather conditions. The microcontrollers on the field machinery coupled with GPS sensors could then make the seed, chemical, or fertilizer applications required by the specific site requirement of the area within each field (DuPont et al., 1999). Harvest equipment using sometimes the same microcontrollers and GPS sensors could record the yield of the field as the harvesters moved across the field generating a spatially-referenced yield map, which then could be used to help plan the next season's prescriptions. The only missing link at this time is the aerial imagery to gauge how the crop is progressing within the season so that mid-course corrections could be applied to maintain or enhance yields.

The primary method for conveying prescriptions from either the farm managers computer where prescriptions were generated or from a third party commercial prescription generator's computer was then and still is to a large extent today the use of PC Cards. The PC Cards would have to be hand-carried to each specific controller located of the farm machinery and inserted into the appropriate slot whereby the application program would then be uploaded and the machinery would carry out the prescription operation.

Today, growers have the controllers and equipment which can carry multiple varieties of seeds and plant these varieties according to soil conditions by soil fertility by soil types or by specific area (AgLeader Technologies, 2008). The GPS signal coupled with the controller and application software determine which variety to plant and how many seeds per row foot to plant. Chemicals applied at planting can be applied at different rates. During the growing season spray equipment can apply as many as six chemicals concurrently by different rates assuming there is no incompatibility between chemicals.

Aerial chemical applications are also beginning to use variable rate technology with GPS sensor technology and high speed controllers changing the chemical application rate as the aircraft speeds across the field.

For many growers the use of PC cards does not cause any problems. However for large farms which have thousands of acres and often are noncontiguous, the PC card limitation is a problem. The farm manager or other skilled personnel have to hand-deliver PC cards to each piece of equipment when precision operations are called for. Many times the farm equipment is in operation and has to be located. All of this takes time away from skilled personnel who could be more productive performing other tasks. If the wrong card is delivered to the wrong piece of equipment, that equipment will have to wait until it receives the right card; again lost productivity or worse the wrong chemical is applied. Major controller vendors have yet to move away from proprietary closed systems to open systems which have telecommunication capability.

Most growers today do not have the training or capability to perform image analysis and application map generation and do not wish to invest in equipment, software or training to be able to do so. They instead will rely on third party prescription generators to perform these operations. Companies exist which can perform all of the image acquisition, image analysis, and application map generation in-house as a turnkey operation. Other companies exist which can acquire the aerial imagery and hand this imagery off to other companies which perform the image analysis and application map generation. There are also some large farm operations which can perform their own image analysis and application map generation.

Another important aspect of prescription farming is ground truthing. The first component of ground truthing is to record actually what was applied or done where. Because of soil conditions, equipment failure, weather, etc, some field operations may not be carried out as required by the application map. In this case, the controller records what is actually done at each time frame and location and this is stored in the controller as an as-applied map. It is vitally important to retrieve this information from the controller and archive the data at the farm headquarters computer because when it comes time to analyze what was done and what was the outcome this information may explain any anomalies.

The second part of ground truthing occurs during the growing season. When imagery is used to generate application maps, the crop and pest management consultants and their helpers have to stay intimately involved in the decision making loop. While imagery has been shown to reduce by as much as 40% of the observations required by consultants to confirm, extend, or compact zones of treatment predicted by the imagery, these still need to be confirmed by boots on the ground. The imagery maps, often in the form of Normalized Difference Vegetative Index (NDVI) maps (Schowengerdt, 1997), can be used to preselect optimum observation or scouting points within fields. The observer navigates to these points using hand-held GPS sensors and confirms the observation or reports discrepancies. These data are then relayed back to image analysis expert who combines the observations and generates the final application map. This operation is another strong requirement for high speed bidirectional data transmission to the farm and within the farm.

In today's farm world most of the events where we are discussing the movement of information or data, it is movement by hand-carrying the data or information to a central site where the data can be loaded in to a base operations computer or downloaded into a microcontroller from a PC Card. The entire US is covered by satellite Internet access which is unsymmetrical in nature. There are fairly high speed downlinks which can reach as high as 1 mbps but the uplink is much slower and peak out at 150 kbps. All of these speeds are relative to the amount of traffic on the shared link and can be much less than the top speeds. As mentioned before, by definition, the great majority of farms are remote and do not have access to ground-based high speed Internet and will not for the immediate and medium term future. The cost of running fiber optic cable is too high and copper will not meet the needs. Telephone modems are not reliable enough to transmit hundreds to thousands of megabyte-sized image files.

CURRENT AND FUTURE TELECOMMUNICATION CAPABILITIES

Current Services

Satellite capabilities have the possibilities of being significantly upgraded, but have the main limitation of being point-to-point and not covering mobile operations without expensive servo equipment to maintain antenna pointing direction. While providing the link to the Internet for the farm operations centers, satellite is not the answer for communicating with farm machinery and pest and crop consultants in the field.

Unlicensed broadband applications have been in use for over 22 years in the US. Most of the applications have become know as WIFI and use the unlicensed 900 MHz and 2.4 GHz bands with some application in the 5 GHz bands, though these are mainly used as links to the 2.4 GHz and 900 MHz sites. Unlicensed WIFI has two major problems for providing the Internet access to farm operations center; these are range and line-of-sight. Rules for use of WIFI were set by the FCC (FCC, 1985). These rules specified the frequencies to be used and the power of the signal to be broadcast. The main problem for 2.4 GHz broadband is the power of the signal is limited to 1 watt so as not to cause problems with adjacent licensed bands above and below 2.400 to 2.483 GHz. The physics of the problem is simple. When you are trying to send high speed data signals, the more data you send at the same power density the shorter the range will be. Further, at 2.4 GHz, the signal propagation is blocked completely by buildings, trees and foliage so line-of-sight is required for relatively long distances. The 5.0 to 5.8 GHz spectrum behaves comparatively the same as the 2.4 GHz spectrum. The 900 MHz spectrum behaves much more favorably than either 2.4 GHz or 5.0 GHz spectrum, but there is less of it. The 900 MHz spectrum ranges from 900 to 928 MHz. Because the frequency is lower, the signal propagates much better at the same 1 watt power density. Non-line-of-sight signals can be reliably transmitted up to a 10 km radius using horizontally polarized antennae. After that, line-of-sight is required. Using line of sight conditions, a central antenna can broadcast reliable signals to a transceiver with a yagi antenna up to 65 km away. Because of the spectrum limitation and FCC rules, the maximum user data rate is about 2 mbps and can be set symmetrical. For more information on WIFI systems and 801.11 standards see Reid (2001), Geir (2002), and Ohrtman et al. (2003).

Because 2.4 GHz WIFI is widely available and costs are very low, a WIFI system can be used to set up high speed data communications link between the farm base of operations and equipment in the field using multiple overlapping WIFI radios broadcasting from the farm's edges inward, much like cellular telephone towers overlap and hand off signals as one travels from cell to cell. Such a system for demonstration purposes was set up at the Good Farm in Noxubee County, Mississippi, USA (McKinion et al., 2004). The base station radio used was an Alvarion BreezeAccess access point (AP) radio with an omni antenna located on the highest structure at the headquarters site. The AP radio provided the link to BreezeAccess subscriber units located at three corners of the contiguous 800 hectare farm as shown in Figure 1. A typical subscriber unit (SU) repeater station is shown in Figure 2. At the repeater station, the BreezeAccess was directly connected to BreezeNet AP radio which then broadcast its signal into the farm property using a sectorial antenna covering a range of almost 4 km. On each farm tractor, combine or picking machine, a BreezeNet subscriber radio combined with an omni antenna mounted on the top of the equipment completed the link from the farm headquarters to the equipment in the field. A user data rate of 2 mbps bidirectional was accomplished.

For farms with noncontiguous fields with treelines in between, a different solution was needed. On a large cotton plantation in the Mississippi Delta in Bolivar County, MS, we tested another demonstration system on Perthshire Farms. Because line-of-sight was an issue with various treelines bordering numerous fields of the 5,000+ hectare farm, a Waverider Model 3001 900 MHz AP radio with an onmi antenna

Figure 1. Map of Good Farm located in Noxubee County, Mississippi, USA showing location of wireless local area network base station (light circle) and three repeater stations (dark outline circles) using 2.4 GHz frequency hopping spread spectrum (FHSS) radios to provide seamless coverage of farm cotton acreage. Vehicle and ground personnel use WIFI radios to complete link for scouting applications with uplink being a FHSS radio

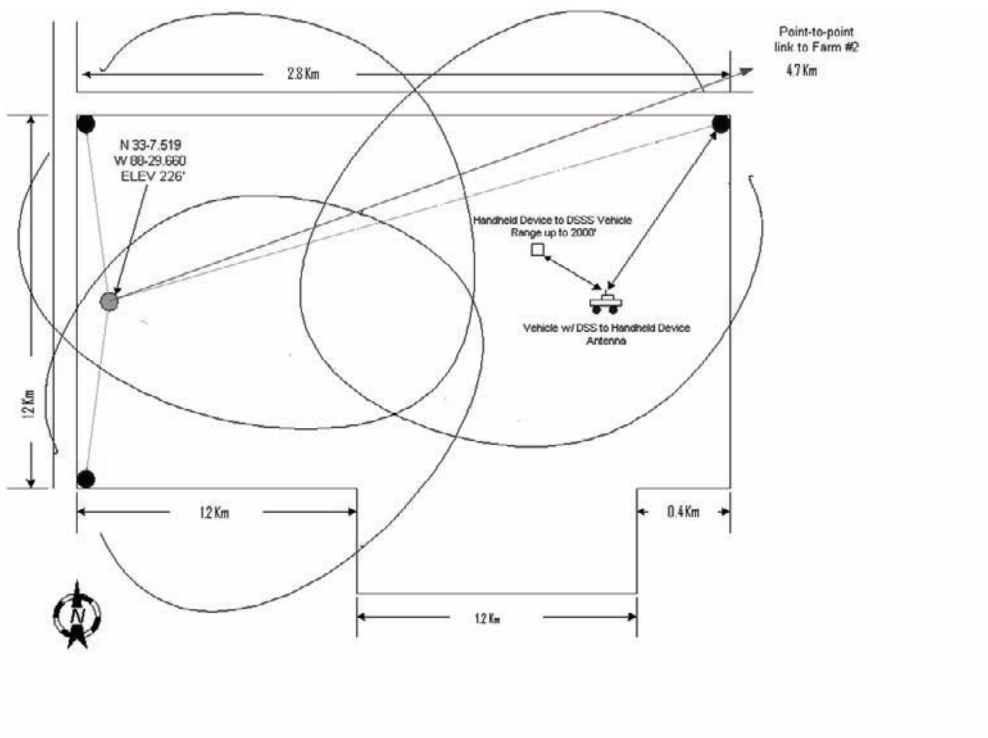
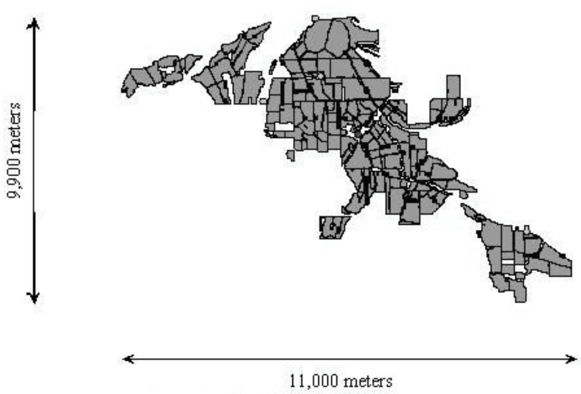


Figure 2. Picture of repeater station with sectorial antenna at top, square panel antenna communicating to base station in the middle of the picture and solar panels for power at the bottom of the tower

mounted on a 60 meter tall guyed tower was used to communicate to eight repeater stations. A Waverider EUM Model 3000 900 MHz subscriber unit completed the link to the farm headquarters with the remainder of the system being the same as the Good Farm. The 900 MHz radio allowed non-line-of-sight operation up to 15 km. In addition to providing high speed bidirectional communication between headquarters and farm machinery, the 900 MHz digital radio system was also used to connect computer systems located at two cotton gins, one close to the headquarters and another 7 km away, allowing rapid communication and system backup from the gin computers to the headquarters computers. Figure 3 shows the geographical layout of Perthshire Farms. The user data rate on this system was 2 mbps up and down.

Figure 3. Map of Perthshire Farms, Bolivar County, Mississippi, USA showing size of area covered



Both of these farm demonstration systems were connected to the Internet using satellite links while very workable but did not allow fast uplinks to transmit large datasets (from hundreds of megabytes to several gigabytes in size) back to image analysts. To address this problem and to explore wide-area networking, a third radio system was established to demonstrate the effectiveness of wide-area telecommunications (McKinion et al., 2007). In Noxubee County, Mississippi on Prairie Point Road approximately 15 km east of Macon, a 100 meter tall microwave tower owned by Teletec Communications, LLC of Columbus, Mississippi was used to place three 120º sectorial antenna configured to broadcast horizontally polarized signals in the 900 MHz band using Motorola Canopy 900 radios. A diagram depicting the sectorial layout of the area of coverage from the 100 m tall tower is shown in Figure 4. Each antenna broadcast at a different frequency for each sector to prevent interference. Non-line-of-sight signal propagation was achieved up to 14 km in radius from the tower. Line-of-sight connections were made up to 60 km from the tower essentially providing coverage to stationary antennae located in Noxubee County, southern Lowndes County, and northern Kemper County in Mississippi, and western Pickens County in Alabama. Internet access to the system was provided by Teletec using high gain point to point dish antenna with the microwave tower being the receive location and Teletec's 125 meter tower being the originating site located in Columbus , Mississippi, 45 km away. Internet access was provided as a full duplex 18 mbps service to the three sector antenna array at Prairie Point. A user data rate of 2 mbps up and

Figure 4. Coverage map indicating the non-line-of sight coverage area advertised by Motorola. 5 km is typical non-line-of sight coverage while we actually were getting 9+ km under non-line-of-sight conditions

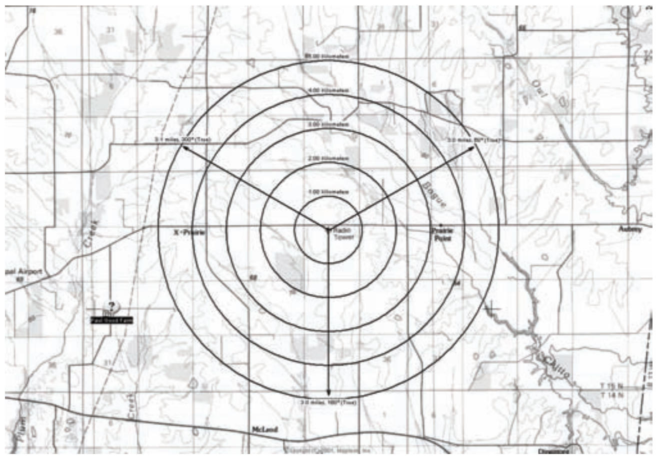


Figure 5. Outdoor antenna mounted at end user site 9 km east of the base station. Antenna and radio are shown on the left with the antenna on top with the radio mounted below. The NEMA 4 box contains the surge protected power strip mounted on the left side interior of the box along with the power converter which supplies 24 VDC power over Ethernet (POE) to the radio, and antenna lightning protector shown mounted on the lower right

down was achieved. A typical user radio link is shown in Figure 5.

Cellular telephone providers recently have extended Internet services called EV-DO and WCDMA which provide medium speed Internet and data communication capabilities which serve mobile applications. Cellular services such as CDMA 1xRTT and GPRS, transmit data at speeds less than 200 kbps in one direction and provide customers with access to mobile data applications such as text messaging, e-mail, and ring tone downloads. Wireless *broadband* networks, on the other hand – such as CDMA 1x EV-DO (EV-DO), Wideband CDMA (WCDMA) with High Speed Downlink Packet Access (HSDPA), and Wi-Fi – transmit data at speeds greater than 200 kbps in at least one direction and provide access to the applications available on the slower networks as well as services that require greater bandwidth, such as video programming, music downloads, and high-resolution. Commercial high speed data communications for farm operations are becoming more available, but current data rates are still fairly low with 800 kbps being the standard. The end user also has to face equipment charges and usage charges. Cellular telephone companies are beginning to extend their data communications systems (EV-DO and WCDMA) to rural areas where they already have cellular telephone service.

One very useful service is the coupling of digital paging systems to critical automated systems in farm applications which send a paging signal and brief message telling the recipient that trouble has occurred and what system has been affected. This system could be greatly helped by being included in a digital communication system so that the recipient could not only recognize the nature of the trouble but take steps remotely to fix the problem or stabilize things until someone could come on site to fix the problem.

Licensed broadband Internet access is a strong possibility which promises high speed Internet access and wide area of coverage. This service is known as WIMAX and has been specifically designed for wireless high speed Internet use. Many vendors now have equipment certified by an independent laboratory in Spain as conforming to the WIMAX standard and interoperable with any other vendors certified equipment. While WIMAX may bring high speed Internet to the rural US, most holders of licensed bands are focusing on populated areas to get the highest return on their investment, and the likelihood of rural applications is still several years away.

WIMAX systems promise much higher data rates and greater areas of coverage than WIFI, EV-DO and WCDMA. Because of these two properties WIMAX should be a less expensive commercial service WIMAX was designed from the ground up to be a wide area data communications system and takes advantage of numerous technical advances such as steerable beam forming to achieve range and non-line-of-sight capability as well as much higher user data rates. WIMAX is beginning to be rolled out but because of economics will be available only in the populated areas first.

Future Services

Higher broadcast power satellites have been proposed using highly directional antenna to service smaller areas of the earth with much higher data rates than currently available (Mir, 2008). For most no-time-critical applications, this would be suitable. But where real-time response is needed, satellites present too great a time delay because of distance involved for the space the signal travels.

The ultimate solution is for every home and business in the US to have access to fiber optic cable, but this service will be a long time coming, if ever, because of the cost involved.

A recent Public Notice published by the Federal Communications Commission (http://wireless.fcc.gov/spectrum/index.htm?job=proceedings_details&proid=369) could have a very large impact on rural America's access to high speed digital communications and the Internet. The FCC has proposed that unused bandwidth in the commercial television channels from channel 2 through channel 51 be allowed to be used as unlicensed spectrum with certain limitations as early as January, 2009. For rural areas in particular and for high population density areas of the country, TV whitespace, unused TV channels, promise very high speed data and Internet access over large distances with signal penetration capability properties which all other broadband licensed and unlicensed spectrum do not possess. Where WIFI signals are blocked by buildings and trees, this is TV spectrum, and we all know that TV signals penetrate buildings and propagate through dense foliage and trees. This spectrum will be available for free usage. Each TV channel uses 6 MHz of spectrum. Assuming a data transmission efficiency of 5 bits/hertz, a single channel could be used to transmit 30 mbps of data. For a rural area like Mississippi or Alabama, the most channels broadcasting in either public or commercial channels in an area is about 6 channels plus channel 39 which is reserved for space usage. This means there would be 43 channels available for high speed digital wireless communication, or in terms of capacity, over 1290 mbps with only using omni antennae and not making use of current spectrum reuse technology at all! Limitations on

Table 1. Comparison of telephone and wireless local area network protocols; GPRS and EV-DO are used alongside cellular telephone networks while the remainder were designed as wireless data network protocols

Protocol	Upload (mbps)	Download (mbps)	Mode
GPRS	0.02	0.08	Full Duplex
EVDO, Rev a	1.80	3.10	Full Duplex
FHSS, 802.11	2.00	2.00	Half Duplex
WI-FI, 802.11b	10.00	10.00	Half Duplex
802.11a,g	54.00	54.00	Half Duplex
WIMAX, 802.16	70.00	70.00	Half Duplex
802.11n	280.00	280.00	Half Duplex

the use of TV whitespace may preclude the use of channels next to broadcasting stations in the lower TV bands, channels 2 thru 13. Data radios will have to have the capability of listening to a channel to ascertain no one is broadcasting on it before the radio attempts to use that channel, otherwise the radio will have to change channels and negotiate the listening process again. All digital radios using TV whitespace will have to be registered with the FCC identifying spectrum to be used and physical location. All radio manufacturers will have to certify and prove to the FCC that their radio equipment will not interfere with public and commercial broadcast stations. These radios will use low power as ordered by the FCC. System builders have said that they can meet all of these requirements. A summary of current wireless technologies is presented in Table 1.

CONCLUSION

Things have advanced greatly from the early use of telecommunications when the only thing being used was the telephone. Next came unlicensed and licensed radio telephone service. Low powered walkie-talkies were useful for small distances but CB radios and licensed two-way radios quickly became the standard for on-farm communication. Some one way communication was used in land leveling operations which used laser technology so that centimeter accuracy could be obtained in elevation establishment in fields for drainage and irrigation flow. When GPS became available for commercial use, another one way communication technology was rapidly adopted. Hand held GPS units were used to mark field boundaries on computer maps which were spatially registered. This confluence of technology, GPS system and digital mapping, allowed precision farming to begin. GPS also allowed the advent of precision guidance of farm machinery, also called auto-steer. GPS was crucial for the development of microcontrollers placed on farm equipment so that as the field equipment proceeded across the field, the GPS sensor told the equipment where it was so that the microcontroller could vary the rate of the application in response to its control program. These last precision agriculture milestones have all used one-way communication. The time has come to close the loop. As commercial and private high speed networks become available to the farm community, all farm data and information traffic can and should become two-way communication. The potential is there for significant savings in manpower and amplification of effort to improve farm productivity. No longer would data cards have to be delivered to controllers and installed by hand. This could all be handled from the farm operations center by trained personnel making best use of their time and effort. Application maps delivered to the wrong piece of equipment could be totally prevented. All equipment could be tracked in real time at the farm operations center. This means that as supplies are being used and applied in the field, operations can track and anticipate delivery of additional supplies to optimize use of farm machinery and personnel.

The availability of systems for use in TV whitespace spectrum could and should have a major impact in rural America. The potential is enormous with bandwidth capacities exceeding that of current state-of-the-art 750 MHz digital cable systems and being equal to or exceeded by only fiber-optic-to-the-curb systems. If this system is put into place in rural America, the digital divide between rural America and urban America will cease to exist.

REFERENCES

AgLeader Technologies (2008). http://www.agleader.com/products.php?Product=pfadvantage. (Verified January 23, 2008)

Daly, P. (1993). Navstar GPS and GLONASS - global satellite navigation systems. *Electronics & Communication Engineering Journal, 5*(6), 349-357.

Dupont, J. K., Willers, J. L., Seal, M. R., & Hood, K. B. (1999). Precision pesticide applications using remote sensing. *Proc. 17-th Biennial Workshop on Color Photography and Videography in Resource Assessment, Reno, NV.*

FCC. (1985). http://www.access.gpo.gov/nara/cfr/waisidx_01/47cfr15_01.html (Verified January 23, 2008)

Geir, J. (2002). *Wireless Lans, Implementing High Performance IEEE802.11 Networks.* Indianoplis, IN: SAMS Publishing.

Geospatial Systems, Inc. (2008). http://www.geospatialsystems.com/imaging-products/ (Verified February, 6, 2008)

Intime, Inc. (2007). http://www.gointime.com/ (Verified January 23, 2008)

Lessiter, F. (2006). Auto-Steer's New Frontier – Precise implement steering. *Farm Eguipment.* http://www.gpsfarm.com/_resources/uploads/files/Farm%20

Equipment%20Sept%202006%20AFTracker.pdf. (Verified January 23, 2008)

McKinion, J. M., Turner, S. B., Willers, J. L., Read, J. J., Jenkins, J. N., & McDade, J. (2004). Wireless technology and satellite internet access for high-speed whole farm connectivity in precision agriculture. *Agricultural Systems, 81*(2004), 201-212.

McKinion, J. M., Willers, J. L., & Jenkins, J. N. (2007). Wide area wireless network (WAWN) for supporting Precision Agriculture. *Proceedings of the 2007 Beltwide Cotton Production Research Conferences, New Orleans, LA*. Unpaginated CDROM.

Mir, R. M. (2008). http://www.cs.wustl.edu/~jain/cis788-97/ftp/satellite_data.pdf. (Verified January 23, 2008)

Raven Industries. (2008). http://www.ravenprecision.com/Manuals/pdf/016-0171/147A.pdf. (Verified January 23, 2008)

Ohrtman, F., & Roeder, K. (2003). *Wi-Fi Handbook: Bulding 802.11b Wireless Networks.* New York, NY: McGrawHill.

Reid, N. P. (2001). *Broadband Fixed Wireless Networks.* New York, NY: McGrawHill/Osborne.

Schowengerdt, R. A. (1997). *Remote Sensing: Models and Methods for Image Processing.* New York, NY: Academic Press.

Walker, W. R. (1989). *Guidelines for Designing and Evaluating Surface Irrigation Systems.* FAO, 1989: ISBN 92-5-102879-6, http://www.fao.org/docrep/T0231E/t0231e01.htm#preface (Verified January 23, 2008)

Willers, J. L., Seal, M. R., & Luttrell, R. G. (1999). Remote Sensing, Line-intercept Sampling for Tarnished Plant Bugs (Heteroptera: Miridae) in Mid-south Cotton. *The Journal of Cotton Science, 3*, 160-170.

KEY TERMS

CDMA: Code division multiple access, a coding methodology for digital radio transmission which improves the amount of information carried by each hertz of frequency. WCDMA is *broadband* CDMA, a further refinement.

DSSS: Direct sequence spread spectrum digital radio transmission methodology which spreads the radio signal over a range of contiguous frequencies to achieve noise immunity.

EV-DO: Evolution Data-Optimized = high-speed mobile data standard used by CDMA=based networks

GIS: Computer based geographic information system technology which records and/or manipulates map data using a map coordinate system to reference map values by pixels.

GPRS: general packet radio service is a standard for wireless communications that allows packets of data, such as e-mail and Web content, to travel across a wireless telephone network and to the Internet.

GPS: The collection of 24 satellites in low earth orbit which transmit precise time information so that receivers can precisely located themselves on the surface of the earth.

Geo-Referenced Map: A computerized map in which each pixel has a value and a geographic information system location reference.

FHSS: Frequency hopping spread spectrum digital radio transmission methodology in which the radio signal is spread over a narrow range of spectrum and the signal hops in a pseudo-random order to other narrow ranges all within a defined overall range to achieve better noise immunity than DSSS, greater range than DSSS and inherent signal security.

Omni and Sectorial Antennae: An omni antenna broadcasts/receives in a 360º pattern completely covering an area while a sectorial antenna broadcasts/receives in a more narrowly defined direction such as 30º, 60º, 90º , or 180º patterns

Precision Agriculture: The practice of planting, applying chemicals, and recording harvest yields based on GIS and GPS information directed by a prescription map using variable rate controllers/recorders on farm machinery.

Prescription Map: A geo-referenced map which contains rate information so that variable rate controllers can apply the appropriate application to the appropriate location using real time GPS sensor information.

TV Whitespace Spectrum: Each television channels occupies 6 MHz of spectrum and in rural Americal there are typically only 6 to 8 channels in use from the set of channel 2 thru channel 51; the unused chanels in this spectrum are called whitespace.

WI-FI: A DSSS radio system based on industry standard IEEE 802.11b which uses the 2.4 GHz signal spectrum, also called wireless fidelity has a range of 10 km line-of-sight and a bandwidth of 11 Mbps.

WIMAX: Wireless metropolitan area network standard IEEE 801.16 which is the new wireless broadband with a range of up to 80km with a bandwith of up to 75Mbps and is the successor to WI-FI.

ENDNOTE

[1] Mention of trade names is for information purposes only and does not imply a recommendation or endorsement by the USDA-ARS.

Section VI
Telecommunications Network Design

Chapter LIV
Network Planning and Dimensioning for Broadband Access to the Internet Regarding Quality of Service Demands

Franz Hartleb
T-Systems, Germany

Gerhard Haßlinger
T-Systems, Germany

Sebastian Kempken
University of Duisburg-Essen, Germany

ABSTRACT

An ongoing challenge in telecommunication is the integration of a variety of services on broadband access platforms at increasing transmission speed. Traditional Internet services like file transfer, email and web browsing, are carried on the same multi service IP platforms with voice, video and television over IP, online gaming, peer-to-peer and grid networking etc. While broadband access is established as a standard equipment for homes, the networking capacities in the access and the backbone are steadily extended to keep pace with higher traffic volumes. Together with the spectrum of services, the traffic mix on the aggregation levels becomes increasingly versatile with different demands for end-to-end transport in terms of throughput, loss and delay sensitivity. The chapter focuses on planning and traffic engineering for link bandwidth and buffers as main resources in communication platforms based on measurement and statistical properties of traffic growth and variability. We summarize quality of service demands of main Internet applications and mechanisms to control and stabilize the performance of ISP network platforms on different time scales. Load thresholds for link dimensioning are derived with regard to quality of service (QoS) demands and the variability in source and aggregated profiles. Finally, link level and network wide traffic engineering is addressed together with load balancing techniques.

INTRODUCTION: GROWTH AND COMPOSITION OF INTERNET TRAFFIC

Information on developing Internet traffic and services does not seem to be available as a global picture. A few periodical reports are provided by standardization bodies and fora, e.g. [5] or in official statistics of countries [1] [16] and some spontaneous sources can be found in market research reports as well as publications from equipment vendors [4] and research institutes [3][15]. The Minnesota Internet Traffic Statistics (MINTS) provides an overview on the home page [20] including links to many relevant sources and measurement data from traffic exchanges in the USA. Odlyzko et al. [15] found Internet traffic to roughly double per year since 1990. Figure 1 shows traffic growth factors based on several previously mentioned sources [1][3][16][20] and Deutsche Telekom's IP platform [17] in the time frame 2001–2007. The steepest curves still follow the trend of doubling per year, whereas moderate cases have annual growth factors in the range 1.5–2. A current white paper of the router manufacturer Cisco [4] gives estimates on the total traffic growth as well as a breakdown into applications. The main conclusions are a forecast of mean annual growth factors of about 1.45 until 2011 mainly driven by several types of video data transfers from peer-to-peer downloads to IP-TV via multicast and video streaming on demand. Within each ISP platform, the product marketing strategy and deployment steps for new transmission and switching technology on the optical and the IP layer have an influence on different growth phases.

Not all factors are clearly predictable and can be planned in advance. When we differentiate the growth due to an increasing subscriber base and the growth caused by in the data rate per subscriber, then the latter is observed to be almost constant on Deutsche Telekoms' IP platform from 2002 – 2007. Although this may be unexpected on first glance, the figures can be interpreted when looking at the fraction of file sharers, who generate the major portion of the Internet traffic since the millennium. File sharing was a driver of the demand for broadband access especially in the early deployment phase of digital subscriber lines (DSL).

In 2002 probably most of the subscribers were running P2P file sharing programs which dominated the traffic demand and smoothed the traffic curve to an almost constant rate over day and night as visible in Figure 2. Meanwhile the population of DSL subscribers in Germany is well beyond 10 million and most of them are using the Internet for multiple purpose. While broadband access entered the mass market, P2P file sharing traffic continued to increase, but the fraction of extensive file sharers was reducing and the traffic profiles per user shifted to applications with lower demand than file sharing. Thus the mean traffic volume per subscriber did not follow the increase in access speed from less than 1 Mb/s in 2002 up to 16 Mb/s or more nowadays.

In the future, the increase in the broadband access population will be relatively smaller than in the last years, whereas the access bandwidth and the traffic generated per user has a large potential for growth when video streaming and television over IP will become feasible and popular. Therefore measurement of the components in the Internet application mix is important for prognosis and planning processes for network expansion.

Odlyzko et al. [15] already pointed out that the traffic mix is changing with a shift from Web browsing applications to peer-to-peer file sharing as the driver of traffic growth since the millennium. Figure 2 shows measurement over four days on a link in Deutsche Telekom's Internet platform in 2003 at the time when file sharing was most dominant. The measurement was obtained on transport layer based on standard or usual TCP/UDP port assignment.

Figure 1. Traffic growth observed on the Internet in different continents

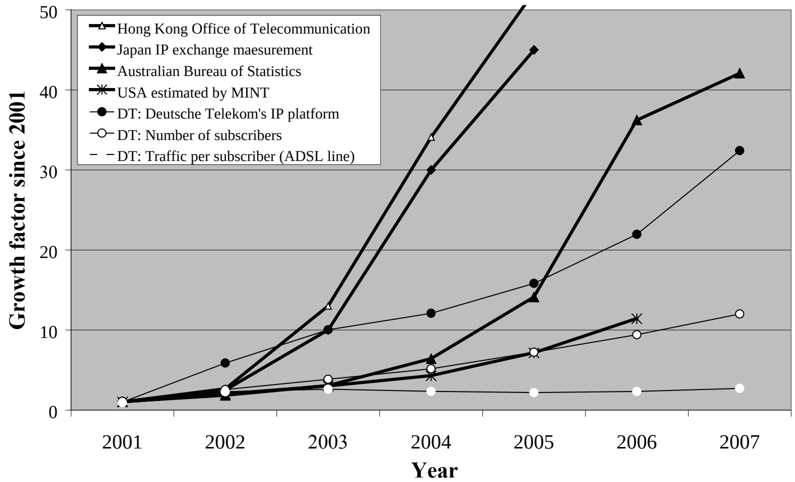
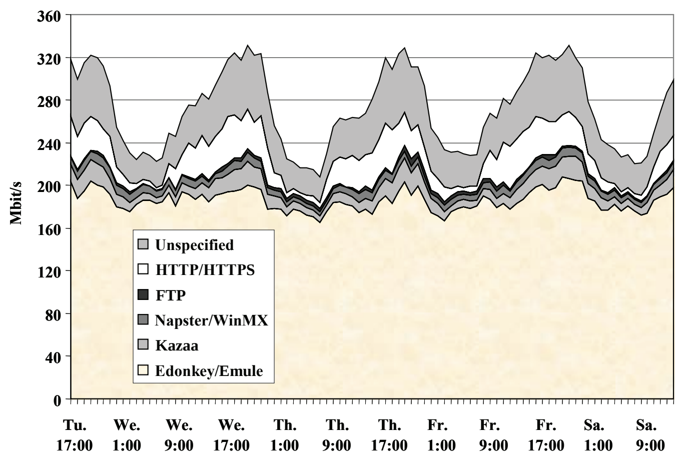


Figure 2. Composition of traffic due to TCP/UDP port measurement



The curves over four days are additive starting from the bottom with peer-to-peer protocols (eDonkey/eMule, Kazaa/FastTrack and Napster/WinMX) followed by classical FTP file transfer and Web browsing via the HTTP protocol and the uppermost portion of all other traffic. The latter includes standard ports with negligible contribution and a larger part of traffic on ports with no clear assignment to an application. Due to the port statistics eDonkey and other peer-to-peer protocols account for most of the traffic. Nevertheless, their variability in the daily profile is essentially lower than for HTTP or FTP traffic. Meanwhile the traffic on unspecified ports has essentially grown, while peer-to-peer traffic is more and more disguising on transport layer using other standard ports or dynamically changing port assignment over a large range, rendering transport layer analysis impossible.

APPLICATION SPECIFIC QUALITY OF SERVICE DEMANDS

Users and application protocols are requesting communication services together with specific expectations of the quality of service (QoS). In a communication network the bandwidth, delay and loss or error rate are decisive transmission parameters with influence on the QoS. The user perceived service quality is attributed as quality of experience (QoE), which obviously depends on the QoS characteristics provided through the network. We refer to UMTS standards [7] for classifying services according to QoS demands as

- Conversational services (e.g., telephony, video conferencing, multi user gaming) for communication between two or more humans in bi-directional or multicast connections with real time demands such that end-to-end delays exceeding 0.25s are intolerable. Picture, voice and audio representations tolerate a bit of transmission errors and packet loss if they are unnoticeable or at least not disturbing.
- Streaming services (e.g., video on demand, Internet television) from a server to humans watching the video or audio stream. The stream is mainly unidirectional or a multi- or broadcast flow to the viewers or listeners while a narrow band channel is maintained in the backward direction for user control. The time sequence of the data stream has to be preserved at the receiver, where a constant transmission delay of some seconds may be tolerable. When the receiver has a large playout buffer, then data can be sent in advance and is stored until its playout time. Again, errors are tolerable if unnoticeable or at least not disturbing.
- Interactive services (e.g., Web browsing, remote access) have a request/response communication pattern between a human and a server. A few seconds of response time are usually tolerable, but immediate response is preferable. Error-free exchange of data is expected and usually enforced by the transport protocol (TCP).
- Background data transfers (e.g., file transfer, file sharing, content distribution) which are initiated by humans but afterwards handled by automated machine-to-machine processes without real time demands since users are not waiting. Traffic for background transfers or downloads may be elastic with varying throughput according to changing network conditions. Background transfers tolerate temporary low throughput or unavailability of the network, that may cause longer transfer times. Again, error-free delivery is expected since a single bit error can corrupt data files in an unpredictable way and would render execution of downloaded software infeasible. Background transfers may even be shifted to low load periods as in case of backups running overnight.

Table 1 gives an overview of the demands of main Internet applications with regard to bandwidth, the service type (conversational, streaming, interactive and background) and the sender-to-receiver relation. The table is not complete and can only characterize some main IP service types of a steadily broadening spectrum with their specific demands. Some application do not fit into the service type classification. The current trend to thin client architectures with remote server control for text editing applications in electronic files via a local network [19] suggests a classification as an interactive service, but can top strict real time demands of conversational services, when the highly delay sensitive movement of a mouse pointer over a screen is included.

CONTROL MECHANISMS ON DIFFERENT TIME SCALES AND IP PROTOCOL LAYERS

Initially, IP networks did not offer much support to guarantee QoS properties and stable network operation, although a fault tolerant routing scheme

Table 1. Service characteristics and quality of service demands

<table>
<tr><th>Services</th><th>Bandwidth demand</th><th>Service type [7]</th><th>Transfer from/to</th></tr>
<tr><td>Telephony,Voice over IP</td><td>Small (< 64kb/s);
continuously available</td><td rowspan="2">Conversational</td><td rowspan="2">User-to-user,
bi-directional
or multicast</td></tr>
<tr><td>Video conference</td><td rowspan="2">Broadband (up to the Gb/s range for HDTV quality);
continuously available</td></tr>
<tr><td>Video on demand, IP TV</td><td>Streaming</td><td>Client/server with narrowband channel from client back to server</td></tr>
<tr><td>Multi user games (eLearning)</td><td>Narrow- (control messages, text) and broadband (graphics); continuously available (broadband for software updates)</td><td>Conversational (& Background for software updates)</td><td>User-to-user, bi-directional or multicast (P2P or client/server for software updates)</td></tr>
<tr><td>Web browsing</td><td>Narrow- and broadband; elastic throughput</td><td>Interactive</td><td>Client/server; bidirectional; request/response</td></tr>
<tr><td>eMail</td><td>Moderate; elastic throughput</td><td rowspan="3">Background</td><td>User-to-user; with multicast via mail server</td></tr>
<tr><td>FTP file transfer</td><td>Moderate or broadband; elastic throughput</td><td>Machine-to-machine</td></tr>
<tr><td>P2P file sharing</td><td>Moderate (music, text files) or broadband (video files); elastic throughput</td><td>User-to-user; distributed transfers in parallel; many-to-many</td></tr>
</table>

was built into the Internet from the beginning. The TCP protocol proved to be another stabilizing factor for overload and error control at least in fixed network platforms even if it cannot easily cope with all scenarios in nowadays next generation networking environments. In pure IP networks, no standard method is available to determine the traffic demand matrix on a platform as a basis for network planning and traffic engineering. Nevertheless, many functions have been added to improve QoS support in the framework of differentiated and integrated services and there is an ongoing standardization by the IETF (Internet engineering task force) for autonomic network operation and management. Multiprotocol label switching (MPLS) is a sub-layer approach introduced by IETF standardization [14] that, among other goals, facilitates traffic measurement and engineering by directing flows on explicit paths through an ISP platform. Meanwhile IP traffic is often directly carried over extended Ethernet access platforms, where QoS and operations support is still at an early stage.

In general, operation and management functions are acting on several time scales. Network elements have to take automatic real time decisions in packet forwarding, including redirection in case of link failures and dropping of packets in case of overload. Link upgrades and topology changes are long term options to respond to changing traffic conditions that can be supported by measurement and reporting tools deployed on the network. Load balancing is another traffic management option by shifting traffic paths to less loaded resources, which has to be updated on time scales from some hours to weeks and after changes in the topology. If overload cannot be avoided, the transmission control protocol TCP executes congestion control by reducing the sender rate of flows that experience packet loss. TCP is an end-to-end protocol mechanism reacting with a delay in the order of round trip times of roughly about a second. Since TCP is not adequate for an increasing portion of real time applications, the IETF standardization is also discussing congestion control for alternative transport layer protocols.

Table 2 summarizes control functions from long term network planning cycles to autonomous control function built into the network elements.

Table 2. Control functions in different time scales

Control Functions	Demands, Input	Goals, Results	Time Scale
Network design, planning and optimization	Traffic demands; technological framework; costs of deployment; reports on developments of utilization, QoS indicators etc.	Topology & dimensioning of network resources; concepts for failure resilience & scalable expansion; load balancing; boundary traffic policies	Months - years
Operations and Management	Basis and status of installed resources; monitoring of network performance measures; alarms on failure conditions	Detection and repair of failures; control of network performance in normal operation and after changes	Minutes - days
Access, Session Control	Current utilization status of network elements; user and service parameters	Admission of user sessions, services, traffic flows; update of current status of resources	Seconds
End-to-End Transmission Control (TCP)	In sequence data reception; acknowledgements; round trip delay estimation; explicit congestion notification (ECN)	Adapting sender rates for a fair share in overload; retransmissions for error-free in sequence data delivery	Seconds
Routing control, Path selection, Failure resilience	Routing-, forwarding tables; link state advertisements; reachability (hello) messages; fast rerouting schemes etc.	Autonomous configuration and discovery of network elements and paths; rerouting; failure resilience	Milliseconds - minutes
Switching	Incoming data packets; addresses, labels; classification flags; buffer status; forwarding scheme (first in first out, differentiated services etc.)	Forwarding of packets based on routing tables; buffering, packet drop; class based and/or fair scheduling policies	Milliseconds

SOURCE TRAFFIC PROFILES OF INTERNET APPLICATIONS

Both, the QoS demands and the traffic profiles have to be taken into account for the dimensioning of network resources including capacities of transmission links, switches, buffers and network servers. High utilization levels can save investment in resources on account of higher risk of QoS degradation in temporary overload and unavailability. In this tradeoff, the impact of utilization levels on QoS aspects is evaluated. Utilization thresholds are determined to trigger resource upgrades in order to meet QoS demands for low delay and data loss rates.

The main aspect that has to be taken into account when setting up utilization thresholds is the variability of traffic. Besides the QoS demands, each application is characterized by a specific source traffic pattern. Main types of source traffic are:

- **Constant bit rate:** The simplest case of a constant bit rate flow may be less relevant in the Internet than various kinds of variable rate flows. Although the nature of video source traffic is highly variable, some video codecs are sending at constant rate to adapt to limited transmission capacity. Peer-to-peer protocols aim at a full exploitation of the access speed of peers, which often leads to a fairly constant up- and download rate over long time. But since most P2P protocols make use of multi source downloads in parallel, the constant rate traffic on the access line is split up into many flows to and from different destinations. File transfers via FTP may also proceed at a constant rate when an access line is fully exploited as the bottleneck on the transmission path. But this depends on network conditions rather than on characteristics of the application.
- **On-off pattern:** An on-off traffic source alternates between pauses and sending activ-

ity at a full rate, which is limited by a codec or the access link speed. When voice over IP includes suppression of silence periods, on- and off-phases are generated according to speech transmission and pauses. Web browsing and other interactive applications often produce on-off traffic exploiting the access speed during on-phases. Pauses as off-phases are partly inserted by the users and partly by the protocols when data units are downloaded from Web sites in successive TCP flows. At a closer look, TCP starts at a low rate which is then steeply increased until a limit is reached usually at the access speed. Therefore TCP flows do not generate a pure on-off pattern and a more precise modeling has to include initial transmit phases with varying throughput [8]. Similar deviations are also relevant for the previous cases with constant bit rate.

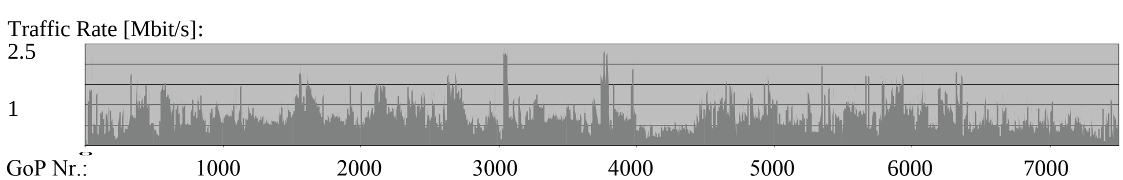
- **Variable rate pattern of various shape:** In general, applications are generating variable rate traffic of different characteristics depending on coding schemes and on the end-to-end interaction between communication partners or client/server systems. Compression and coding schemes are applied to voice, video and data transfers of almost any kind in order to reduce bandwidth requirements. Compression of real time data usually increases the variability of the traffic over time. High variability is also caused by spontaneous and randomly initiated transfers of large data files with arbitrary pauses in between, but is often smoothed due to limited bandwidth available in the access.

When we consider video coding according to current MPEG and H.264 standards [12][18], traffic variability becomes visible on several time scales including

- Periodically changing rates over predefined groups of pictures (GoP), each of which consist of the same sequence of frame types at different compression level (I-Frames: intracoded, i.e. independent of other frames; P-Frames: predicted from previous frames; B-Frames: bi-directional, i.e. predicted from previous and following frames). Each GoP has the same length of usually less than a second,
- Variable intensity of motion during a scene with impact on the redundancy versus coding gain in subsequent pictures that causes short range correlation in the time scale of seconds up to minutes,
- Variability of the compression gain due to scene changes on a longer time scale introducing long range dependency of the traffic rates.

Figure 3 shows the variability of real time video traffic for the example of Jurassic Park. A 1.5 hour trace includes the sizes of 7500 groups of pictures (GoPs) evaluated from data provided on the Web <trace.eas.asu.edu/>.

Figure 3. Variability of real time video traffic



AGGREGATED TRAFFIC: STATISTICAL MULTIPLEXING AND DIMENSIONING FOR GAUSSIAN TRAFFIC

Network links carrying variable rate traffic have to provide sufficient capacity at least for the mean traffic rate and, in addition, to cope with temporary load peaks. Internet traffic variability in different time scales is taken in to account as the basis for dimensioning rules and is decisive for the efficiency of buffers in switching systems.

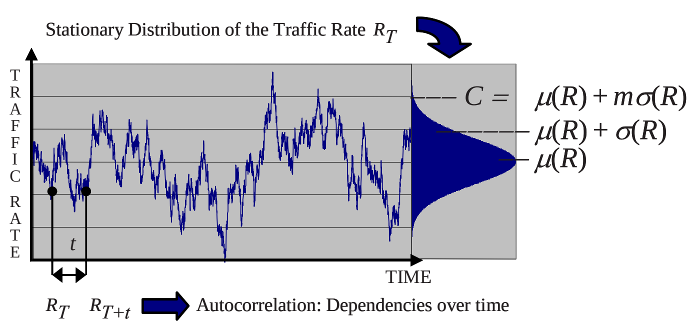
Buffers in routers and switches store data to avoid loss in temporary overload phases on account of forwarding delay. In this way, only short term overload can be compensated until the buffer size is exceeded. Real time applications with bounded delay also impose a corresponding limit on the buffer size. In fact, a fully occupied buffer of usual size introduces a delay of less than a second in order to forward its complete content. Long lasting overload periods or long range dependencies in the traffic rate detract from the efficiency of buffers, as observed for many traffic types e.g. aggregated traffic from Ethernet and DSL access networks [2][10][13] or caused by scene changes in single video streams.

In IP and other packet switching networks, all traffic flows can share network resources with time varying composition of demands for each single flow. Circuit or wavelength switching network layers usually reserve parts of bandwidth exclusively for each connection or flow. Then the peak rate of each flow has to be taken into account for dimensioning, leading to a waste of resources especially for many transmission types with highly variable data rate. Figure 3 demonstrates that the ratio of the mean to the peak traffic rate is small for video streams, which limits the utilization in exclusive per flow reservation schemes.

Instead of looking at single flows, the complete aggregation of all flows on each link is considered in the dimensioning of IP networks. Since the start of transmission phases and pauses of user sessions are fairly independent of each other, the central limit theorem of statistics governs the multiplexing of their traffic flows. The statistical multiplexing effect leads to a smoothed variability of the traffic rate allowing for higher resource utilization.

As another consequence of statistical multiplexing, the traffic rate approaches a Gaussian distribution as depicted on the right of Figure 4. The mean $\mu(R)$ and the standard deviation $\sigma(R)$ or variance $\sigma^2(R)$ are sufficient to determine a Gaussian distribution. In a superposition of independent traffic flows, the mean and the variance of each component are added to obtain both parameters for the total traffic, even if the single flows have different rate distributions. This provides a simple approach to get the main parameters to characterize aggregated traffic behaviour.

Figure 4. Descriptors of traffic variability



$$r_{\text{Loss}} = \int_{x > C} (x - C)\, dF_R(x);$$

$$p_{\text{Loss}} = \frac{r_{\text{Loss}}}{\mu(R)} = \frac{\sigma(R)}{\mu(R)} \left(\varphi(m) - m\,\Phi(m)\right);$$

$$\varphi(m) = e^{(-m^2/2)} / \sqrt{2\pi};$$

$$\Phi(m) = \int_m^{\infty} \varphi(x)\, dx; \quad m = \frac{C - \mu(R)}{\sigma(R)}.$$

DIMENSIONING OF LINK CAPACITIES

Figure 4 suggests that sufficient capacity C for aggregated traffic can be provided by the rule $C = \mu(R) + m\sigma(R)$ accounting for the mean rate $\mu(R)$ and, in addition, a multiple m of the standard deviation. Assuming a Gaussian distribution, the multiple m is determined via the formulas on the right of Figure 4 depending on a predefined demand for the packet loss rate. The analysis does not include buffering and thus represents a worst case scenario where overload leads to loss of a corresponding fraction of the traffic [8][10]. The loss rate is computed by:

The computation of r_{Loss} takes periods into account, where the traffic rate x exceeds the provided capacity C, such that a fraction $x - C$ cannot be forwarded and is lost without buffers. This formula refers to a continuous time or fluid flow model, whereas the forwarding is done per packet. Nevertheless the approach is appropriate, because the size of IP packets is usually smaller than 1500 byte such that thousands of packets are forwarded per second by switches on high speed links. The calculation starts with regard to an arbitrary rate distribution $F_R(x)$, whereas the packet loss probability p_{Loss} is given for a Gaussian rate distribution $\varphi(x)$ with mean $\mu(R)$ and standard deviation $\sigma(R)$.

Table 3 shows typical values of m, which are sufficient to guarantee loss probabilities p_{Loss} in the range 10^{-3} - 10^{-6} depending on the coefficient of variation $\sigma(R)/\mu(R)$. It can be concluded that $C = \mu(R) + 3\sigma(R)$ is sufficient as a dimensioning rule to obtain loss probabilities below 10^{-4}, provided that $\sigma(R) / \mu(R) < 0.25$.

Table 3. Dimensioning for Gaussian traffic: Evaluation of the parameter m and thus the required capacity $C = \mu(R) + m\sigma(R)$

m: $C = \mu(R) + m\,\sigma(R)$	for loss probability demand			
$\sigma(R) / \mu(R) =$	10^{-3}	10^{-4}	10^{-5}	10^{-6}
0.5	**2.50**	**3.17**	**3.76**	**4.30**
0.2	**2.20**	**2.92**	**3.54**	**4.08**
0.1	**1.94**	**2.71**	**3.36**	**3.92**
0.05	**1.66**	**2.50**	**3.17**	**3.76**

As an example, we evaluate the model for traffic flows generated by ADSL subscriber lines with maximum rate $R_{\text{Max}} = 3$ Mbit/s, assuming a mean utilization $p_{\text{On}} = 20\%$ of the access speed during online sessions. Then we obtain for a single on-off traffic source:

$$\mu(R) = 600 \text{ kbit/s} \quad \text{and}$$
$$\sigma(R) \le \mu(R)\sqrt{(1 - p_{On}) / p_{On}} = 2\mu(R).$$

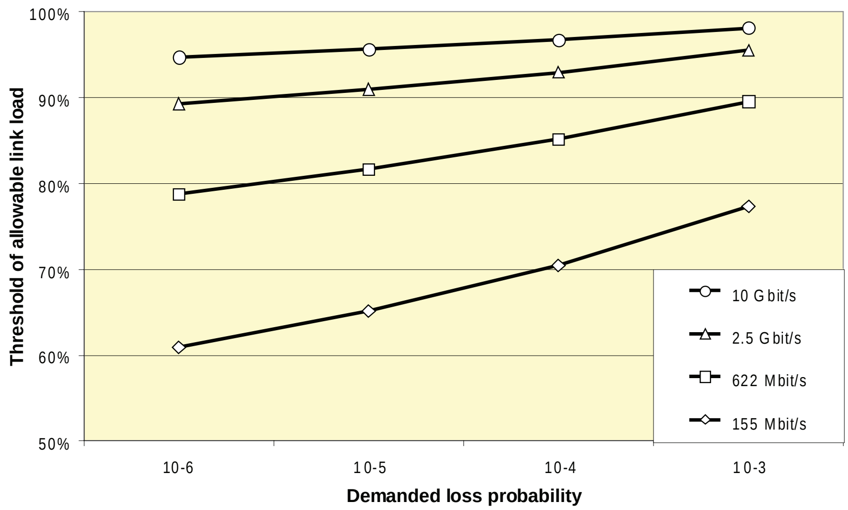
The on-off source model, which alternates between full access speed and transmission pauses, represents a worst case scenario with the maximum variance $\sigma^2(R)$ among all distributions of the access rate with the same mean $\mu(R)$ and rate limitation R_{Max}. Then a multiplexed sum S of M homogeneous ADSL traffic flows with those parameters has mean rate $\mu(S) = M\mu(R)$ and for independent flows the standard deviation is

$$\sigma(S) = \sqrt{M}\sigma(R) \le 2\sqrt{M}\mu(R) \Rightarrow$$
$$\frac{\sigma(S)}{\mu(S)} = \frac{\sigma(R)}{\sqrt{M}\,\mu(R)} \le \frac{2}{\sqrt{M}}.$$

The latter property implies, that the standard deviation $\sigma(S)$ is decreasing with the square root of the number M of independent flows. This characterizes the smoothing effect of statistical multiplexing.

For a link capacity C, the utilization is $\rho = M\,\mu(R) / C$ and we obtain a multiple $m \ge (C - M\mu(R))/(2\sqrt{M}\mu(R))$. In the examples of Figure 5, loss probabilities have been evaluated as a function of the utilization ρ for a set of usual link speeds. Aggregation levels in the access from the DSLAM (Digital subscriber line access module) to the backbone network are connected via STM (synchronous transmission module) links starting from 155Mbit/s (STM-1), 622Mbit/s (STM-4),

Figure 5. Statistical Multiplexing: Loss probabilities obtained for different link speed and load



2.5Gbit/s (STM-16), etc. Assuming that all traffic is generated by ADSL sources with mean rate μ(*R*) = 600 kbit/s, the multiplexing degree includes slightly more than 250 flows on STM-1 links and more than 1000 on STM-4 links, respectively.

It is apparent from those results

- That favourable QoS conditions can be met on backbone links even up to utilization levels of 70% – 90%, but are severely affected as soon as this threshold is exceeded. Therefore appropriate safety margins should be subtracted from the threshold in dimensioning rules,
- That most of the statistical multiplexing gain obviously can be exploited according to the previous analysis without involving buffers.

Nevertheless, the traffic variability in the considered time scale of microbursts is only one of several factors to be included in dimensioning thresholds, where the daily traffic profile as shown in Figure 2 and long term traffic growth processes have to be incorporated leading to further deduction from the allowable utilization level.

In fact, traffic is less homogeneous than assumed in the previous modeling examples. Several source types for ADSL access at different speed as well as traffic entering the network from other platforms, e.g. Ethernet from 10Mb/s to Gb/s speeds have to be taken into account. Inhomogeneous source characteristics detract from the smoothing effect at a considered multiplexing level. On the other hand, there are source types with essentially smoother traffic profile than on-off traffic. Peer-to-peer file sharing applications are expected to down- and upload data in long-lasting transfers without exploiting the full downstream rate.

For many source traffic types, the mean and standard deviation is at least approximately known. Comparing video data files, large differences are encountered in the obtainable compression ratio, which make a common classification and control of real time video transmissions difficult. Nevertheless, it is experienced that an aggregation of about 10 or more independent video streams again leads to an approximately Gaussian shape of the traffic rate distribution.

The independence of traffic flows as a basic precondition of statistical multiplexing is only valid, if synchronization of sources can be avoided, which might arise from TCP flow control in overload situations. Even if the dimensioning of links and forwarding capacities in telecommunication networks can be efficiently done using the considered zero buffer approach, buffers are necessary not only in the end systems on the sender and receiver side but also in switching elements. Multiplexing of a number of asynchronous links requires some small amount of synchronization buffer and a network-wide support of the TCP window mechanism requires buffer for a complete TCP window size for each flow on its transmission path.

NETWORK WIDE TRAFFIC ENGINEERING AND LOAD BALANCING

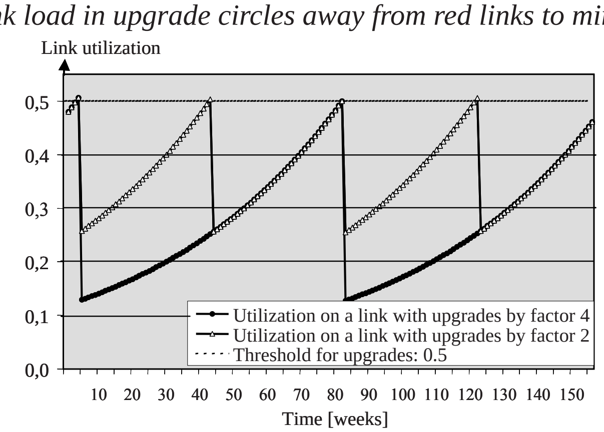
The deployment of multiprotocol label switching (MPLS) [14] on ISP platforms opened new opportunities for network management and planning. While routing in pure IP networks always follows the shortest path, explicit data paths (LSPs: label switched paths) can be set up for traffic flows in MPLS networks providing the required flexibility to enable path designs for traffic engineering purposes. In addition, standard measurement of path statistics makes the traffic demand matrix available as a basis for network planning.

MPLS allows to balance the load among the links of the network. Where shortest routes lead to overload, some demands are shifted to alternative paths in order to balance the traffic flows to make full use of the available capacities. Offline optimization algorithms based on linear programming or simulated annealing are integrated into tools to calculate the most efficient path design [11]. Since quality of service (QoS) properties mainly depend on the load conditions on traversed links, QoS degradations are reduced by adjusted loads or, as a next step, the admissible traffic throughput can be increased while still ensuring a predefined QoS level.

It is important to update the path design at least after a topology change or link upgrade. Without employing network wide traffic engineering, upgrades are triggered separately for each link when a load threshold is approached. This was the usual procedure in pure IP networks for a long time in the past. Meanwhile, IP level traffic engineering is also proposed by manipulating routing weights of links, which has some limitations. Link weights is a less granular and have a more complex impact compared to directing each flow on a dedicated path.

When a link upgrade increases the capacity by some factor, then the load on the link is reduced by the same factor if traffic remains unchanged after the upgrade without redirection of transmission paths. Even if Figure 1 shows

Figure 6. Gaps in the link load in upgrade circles away from red links to minimize overload



that fast and steadily increasing Internet traffic will soon fill new transmission capacities, there are large gaps in the utilization curves after link upgrades depending on the increase factor, which is usually 2 or 4 in the STM hierarchy (STM-1 ≅ 155Mb/s; STM-4 ≅ 622Mb/s; STM-16 ≅ 2.5Gb/s; STM-64 ≅ 10Gb/s; etc.) or may be as high as 10 for Ethernet upgrades. Figure 6 illustrates that those gaps can essentially reduce the achievable link load below a desired threshold level during the complete upgrading cycle assuming a steady exponential increases of the traffic. For upgrades to twice the capacity each time when the threshold is reached, the resulting saw tooth curve reaches only 72% of the threshold in the mean and for upgrades to 4-fold capacity only 54%. This effect essentially contributes to low utilization on Internet platforms.

A traffic engineering solution reoptimizes the path design on the platform after upgrades and can immediately exploit the newly available capacity by shifting traffic path with regard to network wide load balancing. In addition to the gain by adjusting the load level close to a preset threshold, the flexibility of redirecting paths is valuable especially in failure cases. With regard to failures, a set of relevant scenarios including single link breakdowns can be pre-calculated considering the topology without the affected links [11]. If some of the failure scenarios lead to unavoidable overload on the remaining links, this indicates that the topology isn't resistant to failures and has to be enhanced with backup capacity. With pre-computed optimum path designs for failure situations at hand, a fast as possible rerouting for the set of modified paths still imposes a challenge.

Load balancing has to be applied with care, since it may detract from QoS properties in two ways:

- When paths deviate from the shortest possible route and traverse additional hops this has an impact on the delay and reliability.
- Link and node failures may have more serious consequences when the mean load is optimized to a higher level, which leaves less unused capacity available to redirect the traffic.

Regarding longer paths, we experienced that nearly optimum path designs can be found with only few deviations from the shortest paths. Constraints on the number of hops or on link weights can be included in the optimization process in order to prevent paths from violating delay bounds. Full exploitation of the traffic engineering gain may put QoS objectives at risk and a compromise between high resource utilization and QoS demands has to be found.

CONCLUSION

The steady increase of Internet traffic volume demands for an ongoing scalable expansion process for ISP platforms where the load on the links is adjusted in a tradeoff between QoS demands and cost effective high utilization levels. Therefore upper bounds on the link utilization are derived from dimensioning rules regarding traffic profiles and QoS constraints. Since a sharp QoS degradation is expected in excess of a critical link utilization, network planning has to include forecasts to ensure QoS through timely upgrades of the topology. Statistical multiplexing has a smoothing effect on short term variability of aggregated traffic, whereas the daily profiles and long term growth factors are more variable and thus less predictable.

Optimization tools for load balancing and traffic engineering can stabilize the throughput a network wide view, when a fine granular control of resource usage is enabled by directing explicit traffic paths through the network. The flexibility to redistribute imbalance of load in parts of a network is most valuable after traffic shifts and topology changes as well as in the planning process by evaluating and optimizing topology design alternatives.

The broadening spectrum of applications steadily changes the Internet traffic composition and poses new challenges in a trend towards broadband real time services. Despite a tremendous growth over recent years, the bandwidth currently available on ISP platforms is still far from being able to satisfy all demands of broadband applications especially for a widespread usage of video in HDTV quality.

Even if traffic engineering has reached an advanced stage for broadband Internet platforms, mobile and wireless networks require even enhanced methods. New routing schemes are currently under study for unreliable environments e.g. in sensor networks with power saving nodes. The Internet standardization has formed working groups on routing over low power and lossy networks and mobile ad hoc networks addressing this area <www.ietf.org/html.charters/roll-charter.html>, <…/manet-charter.html>. Traffic and network management will develop to cope with the challenges in heterogeneous domains of next generation networks.

REFERENCES

[1] Australian Bureau of Statistics, Internet activity, Latest issue (2007) <www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/Main+Features1Mar%202007?OpenDocument>

[2] H. van den Berg, M. Mandjes, R. van de Meent, A. Pras, F. Roijers and P. Venemans, QoS-aware bandwidth provisioning for IP backbone links, Computer Networks 50 (2006) 631-647

[3] K. Cho, K. Fukuda, H. Esaki, A. Kato, The impact and implications of the growth in residential user-to-user traffic, ACM Sigcomm Conf., Pisa (2006) <www.acm.org/sigs/sigcomm/sigcomm2006>

[4] Cisco Systems, Global IP traffic forecast and methodology, White paper, <http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/net_implementation_white_paper0900aecd806a81aa.pdf>

[5] DSL-Forum, <www.dslforum.org/dslnews/pdfs/YE06_Release.pdf>

[6] EURESCOM, Traffic modelling for the new wireless services (TRAWIS) Eurescom study program P1443 (2004) 1-38 <http://www.eurescom.eu/Public/Projects/P1400-series/P1443/default.asp>

[7] European Telecommunications Standards Institute (ETSI): Universal mobile telecommunication system (UMTS) – QoS concept and architecture, TS 123 107 / 3GPP TS 23.107 (2002) 1-37

[8] F. Hartleb and G. Haßlinger, Comparison of link dimensioning methods for TCP/IP networks, Proc. IEEE GLOBECOM, (2001) 2240-47

[9] G. Haßlinger, Quality-of-service analysis for statistical multiplexing with Gaussian and autoregressive input modeling, Telecommunication Systems 16 (2001) 315-334

[10] G. Haßlinger, J. Mende, R. Geib, T. Beckhaus and F. Hartleb: Measurement and characteristics of aggregated traffic in broadband access networks, Proc. 20. Internat. Teletraffic Congress, Ottawa, Canada, Springer, LNCS 4516 (2007) 998-1010

[11] G. Haßlinger, S. Schnitter and M. Franzke, The Efficiency of Traffic Engineering with Regard to Failure Resilience, Telecommunication Systems Vol. 29/2, Springer (2005) 109-130

[12] S. Kempken and W. Luther, Modeling of H.264 high definition video traffic using discrete time semi-Markov processes, Proc. 20. International Teletraffic Congress, Ottawa, Canada, Springer, LNCS 4516 (2007) 42-53

[13] W. Leland, M.S. Taqqu, W. Willinger and D. Wilson: On the self-similar nature of Ethernet traffic, IEEE/ACM Trans. on Networking 2 (1994) 1-15

[14] Multiprotocol Label Switching, IETF Working group <www.ietf.org/html.charters/mpls-charter.html>

[15] A. Odlyzko, Internet traffic growth: Sources and implications, Proc. SPIE Vol. 5247 (2003) 1-15

[16] Office of the Telecommunications Authority (OFTA) of the Hong Kong Special Administrative Region, Statistics of Internet Traffic Volume (Customer Access via Broadband Networks) (2008) <www.ofta.gov.hk/en/index.html>

[17] W. Schmitz, New telecommunication applications and their impact on network platforms - Future of information technology, Proc. 50th anniversary Informationstechnische Gesellschaft ITG, Frankfurt, Germany (2004) 61-68

[18] P. Seeling, F. Fitzek, and M. Reisslein, Video Traces for Network Performance Evaluation, Springer (2006)

[19] B. Staehle, A. Binzenhöfer D. Schlosser and B. Boder, Quantifying the influence of network conditions on the service quality experienced by a thin client user, Proc. 14. Conf. on Measurement, Modeling and Evaluation of Comp. and Comm. Systems (MMB), Dortmund, Germany (2008) 173-187

[20] University of Minnesota, MINTS: Minnesota Internet traffic studies <http://www.dtc.umn.edu/mints/references.html>

KEY TERMS

Buffers in Communication Networks: Buffers temporary can store data in terminals and switching systems of communication networks. When data is arriving faster than it can be processed then buffering is useful in order to avoid dropping a part of it. On the other hand, buffers in network elements introduce additional delay which limits their efficiency for real time data transfers. Sender buffers help to reduce the variability of source traffic profiles and may transform a variable rate input stream into a constant rate output. In terminals, buffers are necessary to adjust different delays (delay jitter) of received data units for real time service and to hold it available until the play out time as well as for reordering out of sequence packets for transport protocols like TCP.

Quality of Service/Quality of Experience: Quality of Service (QoS) and quality of experience (QoE) measure how much a telecommunication service achieves the expectations of the users. Quality of service introduces performance criteria for data delivery in transmission networks, with regard to throughput, delay, delay jitter and availability as main parameters. Quality of experience directly expresses the view of the user and determines the influence of QoS parameters on the user satisfaction. Methods to evaluate user satisfaction on a scale for the mean opinion score (MOS) are partly standardized e.g. for telephony and partly under research e.g. for video.

Statistical Multiplexing: Statistical multiplexing is performed by switching systems in communication networks that merge data packets from multiple input lines and forward them to multiple outputs in a first come first serve or other scheduling discipline. In this way, many data flows can share capacity on a common transmission path. The aggregation of flows in a multiplexer is governed by statistical laws, such that the entire flow usually shows a smoothed rate variability as compared to single flow components.

Traffic Engineering: Traffic engineering covers all measures to optimize and control traffic flows in a telecommunication network in order to ensure a maximum throughput and a sufficient QoS level. It is a part of the network planning, operation and management process of an ISP.

Concepts for dimensioning, admission control, differentiation of services and failure resilience are included which should ensure a well balanced load level for good performance in normal operation and should keep important services available even in relevant failure scenarios.

Traffic Flow: A traffic flow carries data that is exchanged between the terminals during a communication service. In addition to unidirectional flows from a sender to a receiver, multicast or broadcast flows distribute data from a source to many destinations via splitting points in switching nodes of the network. Traffic flows on the Internet are transferred as a sequence of IP packets with protocol headers being added to the payload data to facilitate routing and delivery through the network. Internet standardization refers to an IP flow as a collection of successive packets with the same IP addresses and transport layer ports for source and destination as well as the same type of service marked in the packet headers. The TCP protocol initiates and terminates a flow by setting up and closing a connection in a handshake with the receiver, whereas UDP spontaneously starts and stops to send packets and flows.

Traffic Profiles: Traffic profiles in communication networks refer to the mean rate and the variability of the traffic rate over time. Applications on the Internet generate their specific source traffic profiles, from which they may be classified as narrow- or broadband, constant bit rate or variable rate. Telephony and voice over IP differs in the traffic characteristic from video streaming and background data transfers, such that properties of source traffic profiles can be used to classify the corresponding application type. Traffic profiles are also measured for aggregated traffic on transmission links.

Chapter LV
Overlay Networks:
New Techniques for Global Service and Network Provisioning

Gerhard Haßlinger
T-Systems, Germany

ABSTRACT

Service and content delivery over the Internet is currently supported by overlay architectures of different types. There is a trend towards distributed computing and service creation in peer-to-peer and grid networks, which are able to overcome performance bottlenecks of client server architectures. Overlays are deployed for single applications or as multi purpose infrastructure for communities with focus on their special demands. Various overlay structures have also developed on lower network layers. The motivation for those overlays is to bridge or extend one networking technology on top of another in order to build common widespread platforms through heterogeneous telecommunication environments. We address such approaches especially within the standardization of Internet protocols (IP), where the main focus is on evolving techniques on higher layers.

1 INTRODUCTION: OVERLAY NETWORKS ON DIFFERENT LAYERS

Overlays on the application layer have developed since the millennium via peer-to-peer protocols where file sharing and voice over IP (Skype) reached world wide popularity. On lower protocol layers, virtual local area networks (VLAN) and virtual private networks (VPN) are well known solution in order to group users communities together by a logical network topology on top of Ethernet or other technologies. The Internet engineering task force (IETF) devoted a considerable part of standardization work to transport mechanisms for support of different platforms over the Internet (IP) and generalized multiprotocol label switching (G-MPLS) networks.

IETF working groups e.g. on pseudo wire emulation edge to edge (PWE3) have defined tunnelling mechanisms to bridge traffic from various platforms over IP including Ethernet and even synchronized transmission for telephony. Other working groups on virtual private networks within different layers (L1-, L2-, L3VPN) extend the basic mechanisms to build overlay networks including virtual private LAN services (VPLS), virtual circuit (VC) emulation etc. based on the previously defined tunnelling mechanisms.

On the application layer, peer-to-peer (P2P) networking overlays became popular in telecommunication through file-sharing applications, which fundamentally changed the traffic profile on the Internet since the millennium. Two main advantages of the peer-to-peer principle as compared to client-server systems are

- A highly efficient distribution of data over the Internet with scalability even for spontaneous requests from millions of users to large files and
- The opportunity to build overlay networks to launch new Internet services around the globe with a minimum of own server and network infrastructure to maintain control.

P2P overlays on the Internet have a profound impact on the business of Internet service providers (ISPs) by making new services available in a short time. Together with the penetration of broadband access, P2P services mark a new phase of Internet evolution with major shifts in technology, service architecture and business models for service providers. Besides file sharing, voice over IP and gaming via P2P systems also have reached populations counting in millions and further applications are expected to make widespread use of pure P2P techniques or hybrid forms combining P2P with server architectures. Functions with a high consumption of system resources and performance requirements in computation power, storage or bandwidth for data exchange can benefit most from P2P and other overlay structures.

Overlay networks generate their own traffic pattern and usually introduce explicit or implicit routing functions e.g. by the selection process of sources for downloads in P2P networks. The data throughput in popular P2P networks is maximized by elaborate techniques including parallel multi-source downloads and incentives to motivate users to make upload capacity available. But transmission paths in P2P overlays are often observed to be suboptimal, since data flows are routed more or less independent of the underlying transport network topology.

As an alternative, content delivery networks (CDNs) have developed from pure content provisioning to support a broadening range of IP services. The advantage of P2P networks can be seen in the opportunity to move operational costs for a service from central servers to the equipment of the users. This facilitates the global launch of new services at low price or for free. On the other hand, security, reliability and control issues are more difficult to handle in a distributed architecture on user equipment. CDNs can enforce full control over the offered content and services at the cost of infrastructure to be installed and maintained within the network. In addition, operational ISPs and telecommunications network providers offer value-added overlay services in a broadband environment. Therefore at least three types of overlays are relevant above the network layer:

- Content distribution networks (CDNs) as global overlays on Internet servers supporting popular Web sites,
- Peer-to-peer applications building overlay networks on the terminals of the users,
- And overlays within the architecture of a network or service provider, utilizing own infrastructure including caches, multi-

/broadcasting or servers in the framework of next generation networks.

These alternative overlays will compete in building a basis for future Internet services, where different approaches may achieve acceptance for one service or another, depending on how efficient the overlay type can meet all relevant demands at lowest possible costs. Among the prerequisites for successful overlays, which decide about which types of overlays will prevail to carry a specific set of applications are

- Efficient search functionality,
- Reliability, replication schemes and security measures,
- Scalability for handling a large number of users, a large amount of data and further network resources,
- Performance and quality of service of the solution in a trade off with costs of the resources,
- Charging principles and incentives for business applications, that are acceptable for all involved parties.

After addressing overlays due to IETF standardization in Section 3, we examine the development and the efficiency of peer-to-peer content distribution in Sections 4 and 5. Cross layer aspects and the potential for path optimization in the underlying network layer are discussed in Section 6. Finally, Section 7 gives a comparison to content delivery networks (CDN).

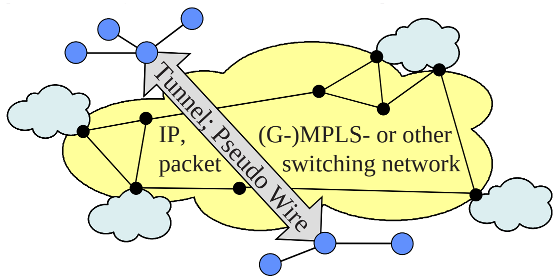
2 INTERNET STANDARDIZATION FOR OVERLAYS ON LOWER LAYERS

Before considering overlays above the network layer, i.e. the IP layer in the Internet, we first address overlays on lower layers combining technologies on the network, the data link and partly on the physical layer. A closer look on work by the Internet engineering task force (IETF) reveals a long history in standardization of overlays around IP. As a first step, a point-to-point tunnel can be set up on a network infrastructure connecting two interfaces attached to networks of a different type at both ends. In this way, different networks protocols can be bridged from access points over the Internet as illustrated in Figure 1. The data of the attached networks is encapsulated and segmented into IP packets when entering the tunnel. After transport over the IP network they are reassembled at the tunnel end to regain the data format valid in the attached networks. Some IETF working groups are mainly focused on tunnelling mechanisms, e.g. the pseudo wire emulation edge to edge (PWE3) working group. The Web site <www.ietf.org/html.charters/wg-dir.html> provides a steadily updated view on IETF working group charters.

In addition, network structures are established as overlays over IP and (G-)MPLS (generalized multiprotocol label switching) networks, which make use of a subset of the core nodes and links. Working groups on virtual private networks (VPNs) have set standards in this area and spread their activity on multiple layers (L1-, L2-, L3VPN on layers 1-3). The number of RFC standardization documents dealing with overlays of various kind is well beyond 250. Usually such work is indicated in a title of the form “protocol or technology X over Y”. Some typical examples are listed in Table 1.

It seems to start with the original role of the Internet of interconnecting Ethernets as the prevalent technology in local area networks. The Internet protocol (IP) is enabled to run over Ethernets and in this way expands into the local domain [13]. This includes specification of the mutual correspondence of IP and Ethernet MAC (medium access control) addresses and gateway functions for converting packet formats. Most data in IP networks is transported in packets of 1500 Byte due to the maximum size to be encapsulated

Figure 1. Tunnels and overlay networks on the Internet



within Ethernet frames.

IP packets are also transported on other link layer technologies for wide area networks, including optical networks [15] (WDM: wavelength division multiplexing, SONET: synchronous optical networks, SDH: synchronous digital hierarchy). Support by corresponding standardization activity is often done in collaboration between the IETF and other fora or standardization bodies for different link layer technologies.

Since the millennium, the IETF defined tunnels and overlays on top of IP and MPLS packet networks as illustrated in Figure 1 [11][12][17]. This becomes challenging when synchronous and connection oriented techniques are transported over connectionless packet networks [11][17], since transmission of IP or MPLS packets is subject to variable delays caused by temporal buffering in routers. Then a dejitter buffer has to compensate for delay variability in gateways from the packet to the synchronous network. Such solutions are desirable for connecting other network types at multiple locations over the Internet, but the efficiency and usability of reversed overlay relationships are not guaranteed in all situations.

In addition, IP-in-IP encapsulation or tunnelling becomes useful in several scenarios, e.g.,

- In mobile networks, when data is first sent to a home agent and then forwarded from the home agent to the current location of the mobile node using a new care-of address or the IP address of a foreign agent of the hosting network,
- For transport of different protocol variants, for security and other purpose or for carrying IPv6 packets over IPv4 networks [4].

Overlay and tunnelling techniques add complexity to the protocol stack and increase the overhead due to additional headers in transmission. There is often a controversial discussion about appropriate use cases and not all options gain wide spread usage as expressed in the document

Table 1. Some IETF standard documents on overlays (RFC: request for comment)

RFC 895 [13]	Standard for the transmission of IP datagrams over experimental Ethernet (IEEE 802) networks
RFC 4717, 4619, 4618, 4448 [12]	Encapsulation for Transport of ATM, Frame Relay, PPP (point-to-point protocol), HDLC (high level data link control), Ethernet over MPLS
RFC 3717 [15]	IP over Optical Networks: A Framework
RFC 4842 [11]	Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Circuit Emulation over Packet (CEP)
RFC 3251 [14]	Electricity over IP
RFC 4817 [21]	Encapsulation of MPLS over Layer 2 Tunneling Protocol
RFC 4906 [12]	Transport of Layer 2 Frames over MPLS
RFC 4798 [4]	Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers
RFC 5087 [17]	Time Division Multiplexing over IP

on electricity over IP [14]. The latter does not refer to powerline or extend power supply integrated in Ethernet, but has deliberately chosen April 1. as editorial date. Referring to a draft version of the SDH/SONET over IP/MPLS work, which was discussed since 2002 and recently entered the proposed standards track [11], the abstract of [14] comments “Readers of the previous work have been observed scratching their heads and muttering: What next?”. In addition to funny suggestions in [14], the current discussion on TCP over UDP <www.ietf.org/proceedings/08mar/slides/mmusic-1.pdf> seems to be another answer, although driven by serious concerns to overcome NAT (network address translation) traversals via UDP, which is often impossible with TCP, while preserving the TCP overload and error control on top.

Other standardization bodies including 3GPP <www.3gpp.org> and ETSI TISPAN <http://portal.etsi.org/Portal> are partly extending IETF work for next generation networks (NGN) by setting up overlays for end-to-end services via different transport media in convergent fixed and mobile networks. The IP multimedia subsystem (IMS) is a well known overlay standard relying on the session initiation protocol (SIP) by the IETF.

3 PEER-TO-PEER NETWORKING: OVERLAYS ON THE USERS' TERMINAL EQUIPMENT

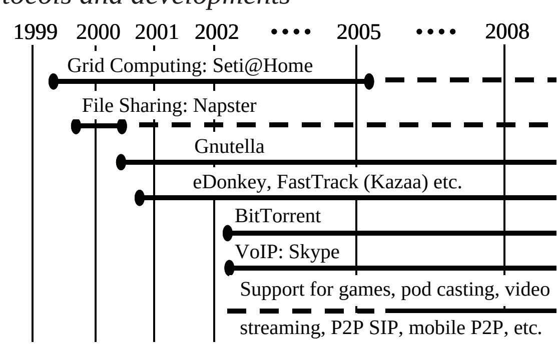
Since the millennium, peer-to-peer (P2P) networking has shown its potential for efficient content distribution and for unleashing idle resources to form scalable and resilient overlays. Peer-to-peer protocols have established popular file sharing and voice over IP (VoIP) applications on a global scale, which offer replicated data distribution schemes and scalable adaptation to the demands at a fraction of the cost of traditional server based approaches. P2P overlays on the end systems of the users give new opportunities for worldwide service provisioning at a minimum of own infrastructure.

In recent years, file sharing often contributes 50% - 80% of the traffic volume on broadband access platforms in Europe and around the world [8]. Although unresolved legal issues still leave an uncertain future for file sharing as a major part of ongoing P2P activity, the spectrum of P2P applications is steadily broadening towards secured research and business applications. The P2P-based Skype solution has become the most popular voice over IP network within less than two years and new P2P approaches are evolving for audio and video streaming, IP-TV, online gaming, cooperative work tools etc. Table 2 shows a brief history of the appearance of some main P2P applications, starting with Seti@Home <http://setiathome.berkeley.edu/> as a network to make computation power available for a scientific project analyzing extraterrestrial signals under central control.

New P2P services are established partly without own server and network infrastructure or in hybrid architectures with a minimum of central server functions in order to maintain control. Ongoing research in P2P networking addresses security functions [7], overlay optimization, search algorithms via distributed hash tables [18] or unstructured search approaches on distributed database systems [20].

Despite of steadily advancing features built into current P2P applications, many of the most

Table 2. Timeline of emerging popular P2P protocols and developments



popular P2P networks seem to be initially launched by small groups of software developers or even by a single person [3], without much involvement of larger players in telecommunication or research institutes. Therefore standardization of peer-to-peer systems is at an early and still premature stage. A working group on peer-to-peer session initiation (P2P-SIP) has been formed by the IETF for locating users and resources for SIP-based communications and presence in a distributed scheme in addition to the server oriented SIP (session initiation protocol) approach.

Application types that are most likely to benefit from peer-to-peer systems are gaming, eLearning, support for enterprises and small communities with extensions into mobile and wireless networks.

- **Online gaming and eLearning:** In both cases, there are similar demands to distribute and update large software packages and to support online activity in multi user games and interactive learning scenarios. The Blizzard gaming platform, e.g., makes use of the BitTorrent file sharing protocol for software downloads <www.blizzard.co.uk/wow/faq/bittorrent.shtml>.
- **Support for small communities and enterprises:** P2P overlays can give access to content and multimedia services according to the specific demands of a community or company including secured and authorized access within closed user groups.
- **Wireless and mobile networks:** Fixed mobile convergence is a challenging environment for P2P systems. Links in mobile and wireless networks have smaller bandwidths and are less reliable. There is a higher churn rate due to changing access conditions of the nodes. Nevertheless, P2P networks have been developed to adapt to unreliable and variable network conditions and provide a promising alternative for ad hoc and self-organizing networks.
- **IP-TV, streaming and multicast services:** Among recent developments, P2P based video distribution is adopted by major broadcasting stations, e.g., in the BBC iPlayer <http://www.bbc.co.uk/iplayer/>, which may be seen a pre-stage of P2P-based Internet television (IP-TV). The BBC iPlayer makes parts of the TV and radio program available for one week after the broadcasting date including digital rights management (DRM) and security enforcement.

The volume of popular unlicensed multimedia content on the Internet is also increasing. There is a variety of providers of open platforms for special purpose with content of various types. Search engine operators also support open content provisioning within their infrastructure. In this environment, peer-to-peer approaches must compete on even grounds with other methods of content distribution. Viability and usability of peer-to-peer networks is measured through technological and economical efficiency and competitiveness.

4 EFFICIENCY OF P2P CONTENT DISTRIBUTION

Analysis of the Delivery Time for Homogeneous Peers and Unique Transfer Times

Next, we investigate P2P content delivery schemes in more detail to get insight into the structures and features required to distribute data volumes over a large community in about the same time as for a single end-to-end data transfer. Content distribution has developed to cover many multimedia applications with prevalent one way transmission from Web browsing to video streaming. We make some simplifying assumptions focusing on a homogeneous overlay network with the same bandwidth B being available at each peer.

We presume unique end-to-end connectivity between peers without considering the underlying network structure in detail. The bandwidth B is bi-directional and can be independently used for up- and downloading in parallel. In fact, ADSL broadband access is prevailingly deployed with asymmetrical up- and downstream speed, but since current P2P protocols enforce a symmetrical traffic in both directions [8], we consider the upstream access capacity as the bottleneck for both directions.

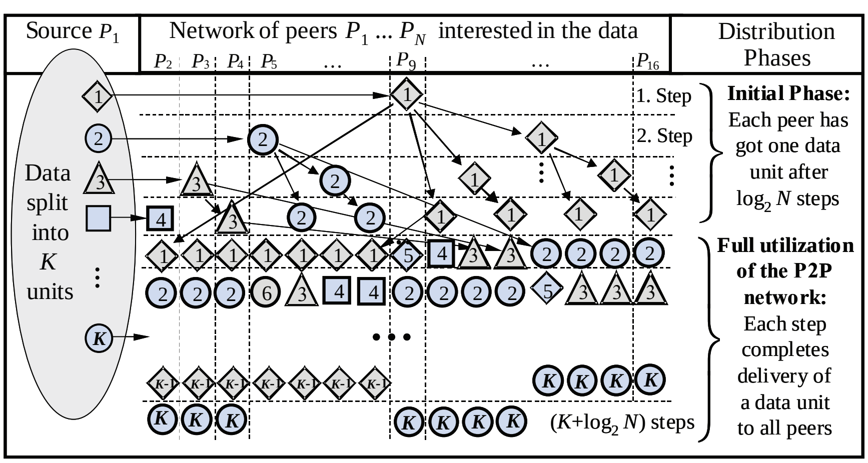
Initially, the content of size C is located on and provided by a source as one of the peers. From the start of the content distribution, we consider N peers who want to get the content and we presume that they build a cooperative P2P subnet to forward the content to each peer without involving others. We assume that the content of size C is subdivided into K data chunks of equal size. The number K of data units is an important parameter for the performance of delivery, since the subdivision into small data portions enables flexible parallel transfers. Each data chunk as well as complete files are uniquely marked by hash identifiers.

Once a peer has completely received a chunk, it can start to forward it to other peers. Upload of received data is enforced as soon as possible by current file sharing protocols [3]. The transmission time for a data chunk is $(C / K + P) / B$ with regard to some overhead P for additional information e.g. a hash identifier being added for unique identification of a chunk. The overhead P may cover connection setup times or other delays occurring between successive chunk transfers.

We assume an unrestricted full mesh of connections between N peers including a source holding the content from the start and the other peers waiting and willing to cooperate in the content distribution. The analysis is done for $N = 2^M$. The source transmits the complete series of chunks one by one choosing different destinations. After the first chunk has been received by a peer, this peer will forward it in parallel to further transfers from the source. As soon as a peer has received at least one chunk, it is assumed to upload a chunk in each step. In a first phase, chunks are always transmitted to peers who still haven't got one. In this way, the number of peers that received a chunk is growing to 2^m after m transmission steps.

Figure 2 illustrates the progress after m chunk transfer times in an example with 2^4 peers. For $N = 2^M$, the first chunk is completely distributed after $M + 1$ steps. At the same time, the second chunk is already available at half of the peers, while the $(M+1)$-th chunk has just been delivered once from the source. In general, the delivery of

Figure 2. Idealized scenario for efficient peer-to-peer content distribution



chunk k is completed in step $M+k$, when again the $(M+k)$-th chunk has just been transferred once. The delivery process can be made periodical from the $2M$-th step onwards, such that the $(j+M)$-th chunk is transmitted over the same peer-to-peer connection as the j-th chunk M steps ahead. There are many appropriate schemes for choosing the destinations for delivery in each step as in Figure 2, which have to ensure that a chunk isn't transferred twice to the same destination.

Thus the content can be forwarded from a source within $M+K$ chunk transfer times to $N=2^M$ other peers in a full mesh P2P network. It can be shown that the same delivery time T is sufficient for $2^{(M-1)} < N \leq 2^M$ and we can conclude

$$T = (\lceil \log_2(N) \rceil + K)(C/K + P)/B = [1 + M/K + (M+K)\,P/C]\,C/B$$

where $(C/K + P)/B$ is the time for one chunk transfer step and $\lceil \log_2(N) \rceil + K = M + K$ is the number of steps.

Usually, we can assume P/C to be small, since a few header or trailer bytes including a hash value should be sufficient as overhead P, whereas the content is often much larger in the MB or GB range. In addition, M can be bounded even in large P2P networks by $M \leq 25$ to include up to $N = 2^{25} > 33$ million peers. Then for $M \leq 25$ and $P/C < 3\times10^{-4}$ we conclude that $K = 250$ chunks are always sufficient to obtain $M/K \leq 0.1$ and $(M+K)\,P/C \leq 0.1$, yielding $T < 1.2\,C/B$. Thus the delivery time T of content over a large P2P network can be kept close to the transfer time C/B of the content to a single destination.

Comparison to Analytic Work Including Delivery via a Tree or Linear Chain

A more detailed performance analysis is given by [1] including heterogeneous peers. In addition to transmission over a full mesh, cases of a delivery over a tree structure and in a linear chain are also investigated. Since no overhead per data chunk is considered in [1], their results are in accordance with the previous analysis for $P = 0$.

Comparing content distribution via chain, tree and full mesh structures for homogeneous peers, the mesh achieves a close to optimum performance under fairly general conditions. The restriction to a tree suffers from the fact, that the upload capacity of the leafs is not utilized, leaving only a fraction $1/h$ of the capacity in the network for the distribution process, and therefore slowing down the delivery by about the same factor. In all cases, a subdivision of the content into small chunks essentially reduces the delay. The performance of the linear chain is essentially affected by additional overhead or delays per chunk and depends on the number of chunks, which has to be optimized with regard to the content size and the length of the chain. For the performance of the tree and mesh distribution, the overhead per chunk is less important and a moderate number of chunks $K = 250$ is always sufficient.

With regard to fault tolerance, an unrestricted P2P mesh offers alternative paths to bypass failed links and nodes, whereas a linear or tree structure offers only a single source to destination path. In fact, only a part of the full mesh connections is required for content distribution. In each transfer step, N parallel connections are utilized. Since the delivery process can be made periodical after $\lceil \log_2(N) \rceil$ steps, only $N \lceil \log_2(N) \rceil$ out of $N(N-1)$ peer-to-peer connections are required, i.e. a partial mesh of no more than $N \lceil \log_2(N) \rceil$ connections already can achieve optimum performance.

Discussion of Content Delivery in Generalized Environment of Realistic Scenarios for Current P2P Protocols

In fact, the previous analysis makes some idealistic assumptions leading to a deterministic and synchronized distribution in a predefined schedule. Popular peer-to-peer networks have to adapt to different delays per data transfer depending on the length of transmission paths between peers in the underlying network. A source decides to upload data chunks based on incoming requests and a local prioritization scheme rather than on a network wide synchronized schedule. Nevertheless, the self-organizing distribution of current P2P protocols is flexible enough to closely approach the optimum performance of the analyzed deterministic scheme.

After the initial phase, the delivery in a mesh structure exhausts almost the complete access bandwidth of homogeneous peers. Therefore only the first M steps could be further improved. In the initial phase, it would be helpful to have a source or peers with larger upload speed involved. A powerful source peer with 2^m-fold bandwidth can distribute 2^m different chunks to 2^m peers in parallel in a single step and thus would skip m–1 steps. But only a small fraction $m/(M+K)$ of time can be saved in this way.

Multiple uploads from the same source in parallel do not improve the performance for homogeneous peers. Simultaneous uploads with bandwidth B / n shared among n chunks of equal size C are all completed after a delay of nC/B. For sequential uploads with bandwidth B, chunks are delivered earlier after C/B, $2C/B$, $3C/B$, …, which is favourable in the starting phase to supply all peers with a chunk and to make their upload capacity available as fast as possible.

However, multiple up- and downloads in parallel are essential in communication between heterogeneous peers equipped with different access speeds. When many peers have a unique basic access speed and there are some peers with higher, e.g. n-fold speed, then it is easy to integrate them by subdividing their speed into n up- and downloads at unique speed in parallel. With a single transmission per peer, it would be difficult to fully utilize peers at different bandwidths due to the restrictions for transmitting between pairs of almost equal speed.

An extension of the performance bounds and estimations to heterogeneous P2P networks is more complex, but surely relevant, since different access speeds are offered for residential users and some very powerful hosts are often involved in P2P networks. Although multi source downloads are appropriate to adapt to the heterogeneity, the delivery time for content of size C starting from a source with bandwidth B_{Source}, is obviously limited by C / B_{Source} as a bottleneck at the source. When the data is available in the P2P network and distributed over many peers, then the access speed of most powerful hosts among the peers can be exploited.

More generally, we may distinguish several classes of peers at different access speed but with the same speed within each class. Then a class of high speed peers including the source can be viewed as a homogeneous subnet, whose delivery time is determined by the access speed of this class, even if the complete P2P network includes many peers of lower speed. In addition, the peers with lower access speed can profit from faster peers, especially in case of asymmetrical access (ADSL).

On the whole, bounds on the delivery time to all peers can be set up in terms of the mean upload speed per peer, since the upstream is assumed as the bottleneck in asymmetrical access, where classes of peers with higher speed may finish a download earlier than peers with smaller access bandwidth. From the view of each peer, the access speed of the peer can impose a bottleneck, as well as the upload speed of the source, if the content is only available from a source and not yet distributed over the network.

Involving more peers from outside of the subnet also wouldn't improve the main distribution phase, since the bandwidth of the peers is already exhausted by the distribution scheme as suggested by Figure 2 within the subnet of all peers interested in the content. For asymmetrical ADSL access with a ratio R of download ÷ upload capacity the situation is different, since then the download capacity can only be exploited by *R* peers uploading in parallel to the same destination.

Solutions for efficient content distribution apply in a similar way to non-real time as well as real-time applications, provided that real time data, e.g. a video stream,+ is again subdivided into data chunks of suitable size.

5 TRANSPORT PATHS AND CROSS LAYER ASPECTS

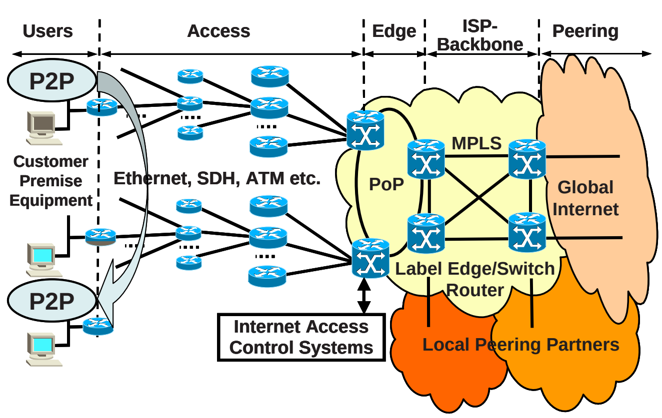
In the previous section, we focused on the application layer, where P2P overlays are shown to be highly efficient for content distribution and in providing a welcome opportunity to launch services at low infrastructure and development costs. However, the advantage of becoming independent from the transport and network layer is achieved on account of some overhead for an additional protocol layer.

P2P protocols establish their own routing schemes by selecting download sources via an inherent search scheme, including "hello" mechanisms to verify the availability of peers. When comparing the message exchange for downloads via the eDonkey and BitTorrent protocol, we experience a messaging overhead of 10-20% considering the ratio of the complete transferred data volume during a download to the size of the file. For small files there is more overhead. The first version of the Gnutella peer-to-peer network, that partly replaced Napster after shutdown, wasn't scalable because of flooding many search messages through large parts of the network [10].

Another major performance issue is the length of paths for data transfers, i.e. the hop distance from source to destination. In P2P networks, frequently referenced data is replicated manifold according to the demands and becomes available from many sources. So, there is a potential to optimize the transfer paths by preferring the nearest sources. In this way, transmission delays and thus the user perceived QoS can be improved as well as the resource usage of network providers, avoiding unnecessary long paths through the backbone or on expensive peering and intercontinental links. Figure 3 illustrates a usual topology of broadband access networks, where tree-shaped access areas are attached to edge routers of the backbone at points of presence (PoPs).

Considering large provider networks serving millions of subscribers, it can be expected that a majority of the data of a global file sharing network can already be found to be replicated on the same ISP platform and often in the same access region of the P2P downloader. This especially holds for highly popular and referenced data, since the major

Figure 3. Usual structure of broadband access platforms



portion of downloads is addressed to a small set of the currently most popular files. But peer-to-peer networks have no direct knowledge of the underlying IP network structure. The fact that the response time and availability of the peer nodes are related to the length of transmission paths on the network layer gives some preference for shorter paths. But this effect is less relevant than structures in the overlay, which are imposed through server or super node locations or communities separated by language and social environments.

The distribution of source locations offered by the eDonkey network for downloading German and English content has been investigated from destinations in Germany with results shown in Figure 4. It is not surprising, that 78% of the sources for German content were also located in Germany, whereas most (83%) of the English content is transferred from peers in other countries, which puts load on backbone and peering links of the network providers. Similar experience for locality of sources due to language has been made for file-sharing in France [9].

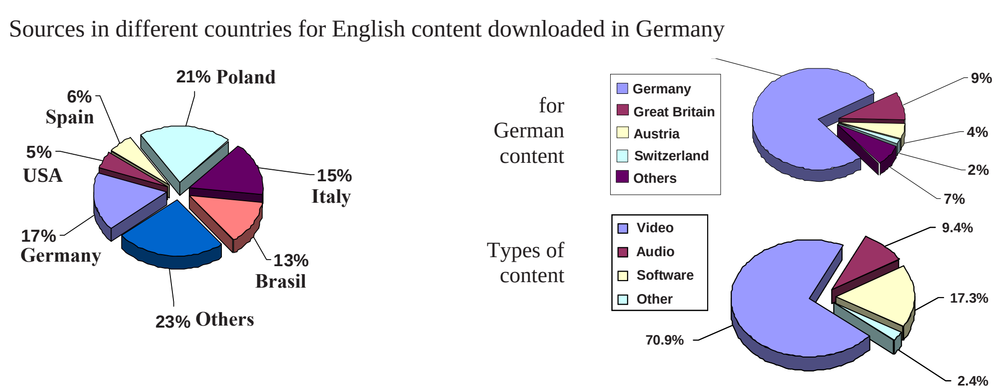
Especially the figures for English content downloaded in Germany demonstrate the global distribution of source locations, resulting in long download paths over intercontinental links and causing long transmission delays. Thus network providers as well as the users and thus the P2P protocols should be interested in preferring peers close to the destination. Since the location of peers can be estimated from their IP address at least with regard to their corresponding autonomous system and network operator, it is feasible to do better. This could be achieved by improved source selection mechanisms in the P2P protocol or by using caches for serving downloads in the proximity.

There are several proposals for cooperation between network providers and P2P protocols for file sharing [1][5][6][22]. Cooperative approaches [1][22] would have the highest potential for efficient solutions by selecting appropriate sources and including caches or even CDN-like architectures [5][6] to support local content distribution. But file sharers, and thus the protocols they use, usually try to avoid that their IP addresses are looked up and processed due to special policies of third parties, even for the goal of optimizing their performance.

6 CONTENT DELIVERY NETWORKS: OVERLAYS EMBEDDED IN THE INTERNET STRUCTURE

Content delivery or distribution networks (CDNs) provide an alternative overlay structure within the network layer and thus within platforms of ISPs.

Figure 4. Sources for downloads to German sites in the eDonkey network and types of content [16]



After a user contacts a Web site, which is supported within the CDN network, his request is redirected via a CDN overlay to a server in the near of the user. A performance study on Akamai's CDN [19] reveals that a large hierarchical server farm is efficiently utilized to shorten the transmission paths. The solution includes frequent updating of the best path to a user based on measurement of the network and server performance. A backup path from another server is also prepared for failure resilience. In this way, CDNs offer advantages over P2P networks for optimized reliability and low delay as main QoS measures. Short transmission paths to globally distributed Web sites reduce traffic in the backbone and on expensive transatlantic links of network providers.

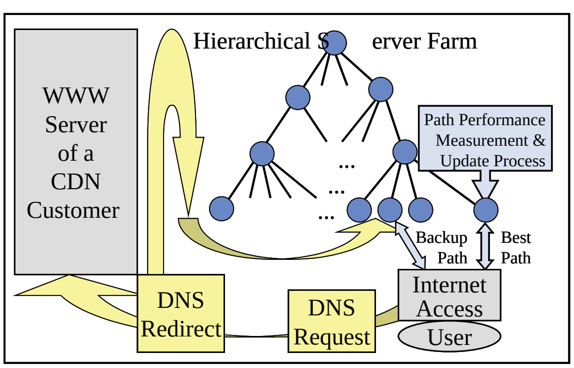
Another advantage of content delivery networks is that content is stored in a controlled way on provider infrastructure whereas it is distributed over customers' and users' equipment in P2P networks. Although a secured and controlled access is also enabled for P2P using strong encryption methods, some content providers may still have more trust in securing valuable content on their own infrastructure. Search engines and the majority of popular Web sites have organized CDN support on their own or with the help of CDN providers like Akamai.

Web caches provide another opportunity to optimize traffic paths. In principle, caches are efficient for P2P data since a high hit rate can be expected and problems of outdated data as in classical Web caches are avoided because the P2P content is uniquely identified by hash values.

Since October 2004, the eDonkey P2P network offers an option to use Web caches of network providers by disguising P2P downloads as usual HTTP Web browsing requests. Indeed, eDonkey downloads can be observed to partly originate from caches, but only to a small extend of less than 10% being indicated in the eDonkey performance statistics [16]. From the network provider's view, the caching option for support of P2P networks remains ambiguous. While transfers from the cache take load away from the backbone and reduce off-net traffic, other downloads may continue without caching and may generate a higher total traffic load. The problem of also supporting the distribution of copyright infringing or other illegal content via caches or other techniques for efficient transport is known since the Internet became popular. But it has attracted much more attention and is pursued by various counteractions since file sharing networks have been made responsible for decreasing revenues for content providers in the music and film industry.

Last not least, network providers have opportunities to build content distribution architectures using their own infrastructure e.g. via multicast for support of broadcast services. Usually this does not lead to global access but is often appropriate for regional services e.g. for IP-TV within a country or lingual community.

Figure 5. Transfer paths via content distribution networks



7 SUMMARY AND CONCLUSIONS

Overlays of various types are used to bridge heterogeneous networks to enable a seamless end-to-end service delivery. This becomes even more relevant in the current fixed/mobile convergence phase integrating wireless and mobile communication on different technologies with Internet access. A major advantage of overlays is their ability to provide new services or extend existing services on top of an underlying infrastructure. Internet standardization by the IETF has made extensive use of tunnelling and overlay techniques.

On the application layer, peer-to-peer networks establish overlays on terminal equipment of the users, offering global services with a minimum of own network infrastructure. Scalable and highly efficient delivery of large volumes of content to a large community of subscribers is characteristic for peer-to-peer networking.

Content delivery networks (CDN) usually do a good job in optimizing transmission paths and corresponding delay by choosing the nearest server in their platform to contact each user. Network providers include overlay traffic in the planning and upgrade process of the network topology including mechanisms for load balancing. But in this way, inefficient routing on higher layers can only be partly reduced by optimization on lower layers.

Overlays add to the number of layers in the communication protocol stack. As a consequence, they increase overhead and cross-layer inefficiency by repeating similar functions on several layers or, as a worst case, by employing functions on different layers, which are jarring with each other. Therefore overlays should be set up with implicit or explicit awareness of the underlying network structure. On the other hand, it is a challenging design question for future Internet architecture to avoid complex layering structure and at the same time to keep the flexibility to respond to new requirements and technical opportunities as the main motivation for the appearance of overlays of various type.

REFERENCES

[1] V. Aggrawal, A. Feldmann and C. Schneideler, Can ISPs and P2P users cooperate for improved performance? ACM SIGCOMM computer communications review 37/3 (2007) 31-40

[2] E. Biersack, D. Carra, R. Lo Cigno, P. Rodriguez and P. Felber, Overlay architectures for file distribution: Fundamental analysis for homogeneous & heterogeneous cases, Computer Networks 51 (2007) 901-917

[3] B. Cohen, Incentives build robustness in BitTorrent, <http://bitconjurer.org/BitTorrent/bittorrentecon .pdf> (2003)

[4] J. De Clercq, D. Ooms, S. Prevost and F. Le Faucheur, Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers, IETF, RFC 4798 (2007)

[5] M. Faycal and A. Serhrouchni, CAP: A context-aware P2P system, Proc. OTM'07, Springer LNCS 4806 (2007) 950-959

[6] R. Ferreira, A. Gramma and S. Jagannathan, Pletora: An efficient wide area storage system, Proc. HiCP'04, Springer LNCS 3296 (2004) 252-261

[7] L. Gaspary, M. Barcellos, A. Detsch, R. Antunes, Flexible security in peer-to-peer applications: Enabling new opportunities beyond file sharing, Computer Networks 51 (2007) 4797-4815

[8] G. Hasslinger, ISP Platforms under a heavy peer-to-peer workload, Proc. Peer-to-Peer Systems and Applications, Eds.: R. Steinmetz and K. Wehrle, Springer LNCS 3485 (2005) 369-382

[9] G. Hasslinger, F. Guillemin, J. Ferreira and U. Halldorsson, The impact of peer-to-peer networking on network operators and Internet service providers, Eurescom study report P1553 (2005) <www.eurescom.eu/public/projects/P1500-series/p1553/>

[10] D. Ilie, A. Popescu, Statistical models for Gnutella signaling traffic, Computer Networks 51 (2007) 4816-4835

[11] A. Malis, P. Pate, R. Cohen, Ed., D. Zelig, Synchronous Optical Network/Synchronous Digi-

tal Hierarchy (SONET/SDH) Circuit Emulation over Packet (CEP), IETF, RFC 4842 (2007)

[12] L. Martini et al., Encapsulation for Transport of ATM, Frame Relay, PPP, HDLC, Ethernet over MPLS, RFC 4717, 4619, 4618, 4448, 4906 (2006)

[13] J. Postel, Standard for the transmission of IP datagrams over experimental Ethernet (IEEE 802) networks, IETF, RFC 895, 1042 (2005)

[14] B. Rajagopalan, Electricity over IP, IETF, RFC 3251 (1.4.2002)

[15] B. Rajagopalan, J. Luciani and D. Awduche, IP over Optical Networks: A Framework, IETF, RFC 3717 (2004)

[16] J. Schroeder-Bernhardi, Analysis of the communication and traffic in P2P networks including Web caches, Diploma Thesis, Report KOM-D-260, Darmstadt Univ. of Tech. (2006)

[17] Y(J). Stein, R. Shashoua, R. Insler, M. Anavi, Time Division Multiplexing over IP, IETF, RFC 5087 (2007)

[18] I. Stoica et al., Chord: A scalable P2P lookup protocol, IEEE/ACM Trans. on Networking Vol. 11 (2003) 17-32

[19] A.-J. Su, D.R. Choffnes, A. Kuzmanovic and F.E. Bustamante, Drafting behind Akamai, ACM SIGCOMM, Pisa, Italy (2006)

[20] W. Terpstra, J. Kangasharju, C. Leng and A. Buchmann, BubbleStorm: Resilient, probabilistic and exhaustive P2P search, Proc. ACM SIGCOMM, Kyoto, Japan (2007) 49-60

[21] M. Townsley, C. Pignataro, S. Wainner, T. Seely and J. Young, Encapsulation of MPLS over Layer 2 Tunneling Protocol Version 3, IETF, RFC 4817 (2007)

[22] H. Xie, A. Krishnamurthy, A. Silberschatz and Y. Yang, P4P: Explicit communications for cooperative control between P2P and network providers, <www.dcia.info/documents/P4P_Overview.pdf>

KEY TERMS

Cache: Caches provide storage within the network infrastructure of Internet service providers as well as for support of servers. Data transfers from remote sources can partly be abbreviated from a cache in the near of the destination, thus enabling optimized transfer paths in a similar way as done by CDNs.

Content Delivery (or Distribution) Network (CDN): Content delivery networks provide support of popular Web sites or search engines in the Internet. They employ an overlay consisting of globally distributed server farms, which enables to optimize data transfer paths and transmission delays by choosing a server close to the location of the requesting user.

Internet Engineering Task Force (IETF): The main technical standardization body for the Internet <www.ietf.org>. The IETF is a non-member organization open for individual volunteers to contribute to Internet protocol development. The IETF in organized in more than hundred working groups on current topics, holds three annual meetings with most of the work being available in Internet drafts and mailing list discussions. The IETF has conducted a lot of basic work for overlays in telecommunication on and around the network layer.

Overlays: A popular method to put new services or technologies on top of an existing telecommunication network infrastructure. Peer-to-peer networks are an example of an application layer overlay. The IETF and other standardization bodies have defined overlays on lower layers for many purposes between packet networks (IP, MPLS), Ethernet and synchronized networks for voice transmission. In this way, a standard networking solution can be carried and thus extended over other networks or combined with them in a heterogeneous networking environment. Overlays add new layers to a communication architecture.

Peer-to-Peer (P2P) Networking: An overlay technique on communication networks, which is implemented in the terminal equipment of users with Internet access. Since the millennium, many applications became popular on P2P networks starting with file sharing, voice over IP solutions.

Tunnel: Transmission through a tunnel establishes a point-to-point connection to carry data of one transmission protocol through a networks using different protocols and/or technologies.

Virtual Private Network (VPN): Virtual private networks provide overlays on the network layer in order to connect separated networks at different locations via a telecommunication network. The IETF has developed a standard suite for VPNs on the Internet. The operator of a VPN can have limited visibility of the underlying network structure of a service provider supporting the VPN.

Chapter LVI
Optimising P2P Networks by Adaptive Overlay Construction

James Salter
University of Surrey, UK

Nick Antonopoulos
University of Surrey, UK

ABSTRACT

Distributed Hash Tables (DHTs) have been used in Peer-to-Peer networks to provide query lookups in typically O(log n) messages whilst requiring maintenance of only small amounts of routing state. We propose ROME, a layer which runs on top of the Chord DHT to provide control over network size through monitoring of node workload and propose the use of processes to control the addition or removal of nodes from the network. We show that this technique can reduce further the hop counts in networks where available node capacity exceeds workload, without the need to modify any processes of the underlying Chord protocol.

BACKGROUND

With many media reports of increased piracy (Digital Spy, 2005) (one file-sharing server recently shut down indexed 170 million pirated files (MPA, 2006)), free-for-all access to illegal materials (BBC News, 2005) and promises of finding answers to previously incomputable problems emerging as a result of them (Anderson, 2002;Bohannon, 2005), Peer-to-Peer networks have become a hot topic not only of interest to the scientific community, but to a wider audience too. Of course, not all coverage has been accurate, but P2P networks have nonetheless attracted much attention.

As an emerging technology, there are many opportunities for research work that will provide real benefits for the future of Peer-to-Peer networks. In this chapter we look at the area of resource discovery. Since large collections of resources are distributed throughout such networks, searching and discovering the whereabouts of them becomes a vital function.

Traditionally, networks were based on client/server architectures. Client workstations would be connected to a central server which would handle all requests. In this environment, resources (for example files or databases) are held on the server, so client-to-client communication is rare or non-existent.

Client/server architectures allow for requests for resources to be resolved efficiently. Each client is directly connected to the server which means, as it is the central resource repository, a single hop over the network is all that is required to send a request. On the other hand, having a central server leads to a single point-of-failure. Whilst communication is still possible if one of the clients fails, if the server is removed from the network then none of the machines are able to access resources.

Largely due to success of the World Wide Web, we often view the Internet as a client/server environment. Clients connect via Web browsers to Web servers that return content in response to requests. However, the original ARPANET was conceived to share computing resources throughout the United States. Machines were connected as equal peers. As the Internet grew it took on a client/server-like architecture. Recently though, the trend has begun to reverse as a new breed of peer-to-peer network has emerged (Minar, 2001).

Pure peer-to-peer (P2P) architectures tend toward the opposite of client/server architectures. There are no central servers. Instead, machines (often referred to as nodes) are directly connected to several others and have the potential to act as both clients and servers. Resources are typically hosted on nodes at the edge of the network, rather than on dedicated servers.

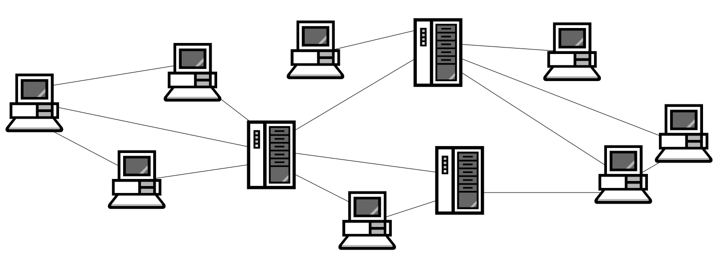
A pure P2P architecture is more robust than a client/server architecture because the need for a central server, and thus the single point-of-failure, is removed, although this comes at a cost. Resources are spread throughout the network so it is no longer guaranteed that a requestor will be directly connected to a provider, meaning it may take several hops over the network to find the required resource.

In practice, it is rare for a network to have a pure P2P architecture. Many are hybrids (Figure 1) positioned in-between pure P2P and client/server. Not all nodes in hybrid P2P architectures have equal standing within the network like in pure P2P. Some nodes take on additional responsibilities, becoming group managers or routers, which leads to more organisation within the network. Typically they hold more information about the network than a standard node, allowing for messages to be routed in fewer hops. Our discussion of P2P networks will treat them as high level networks running at the application layer. These networks are overlaid on top of the lower level networking layers, hence the term *network overlay.*

RESEARCH QUESTION

Previously, research into creating efficient resource discovery schemes for use in P2P environ-

Figure 1. Hybrid Peer-to-Peer Architecture



ments has succeeded in developing models based on Distributed Hash Tables (DHTs) that distribute an index of resources across network nodes and are capable of resolving queries in typically O(log *n*) messages, where *n* is the number of nodes in the network (Zhao, 2003).

Reducing the volume of messages required has concentrated on optimisations to the distribution of the index and the amount and type of routing information stored. However, since the number of messages is dependent on *n*, the number of nodes in the network, we believe that by creating a scheme to control how many nodes are members of the network, more efficient resource discovery can be provided. This leads us to our research question:

Can we provide efficient, fault-tolerant resource discovery in peer-to-peer networks by controlling construction of the network overlay?

Hosting the entire index on a single node would indeed allow for efficient discovery of resources (using O(1) messages) but has a single point-of-failure and, if the node is not well-chosen or the number of queries is very high, could cause the node to become overloaded. Simply saying that the network size should be minimal (i.e. a single node) is therefore an over-simplification, so consequently it is important to consider the fault-tolerance of the model in addition to its efficiency.

Control of the number of nodes entering the network is not the only parameter that will affect the efficiency and fault-tolerance of the model. For example, a network constructed of a small number of low-quality nodes may not be any more efficient or fault tolerant than a larger network of higher quality nodes, because queries have a higher tendency to fail.

Efficiency not only relies on a small number of messages required to resolve a query. We must take into account the number of messages required to maintain the network. For example, a model would not be efficient if, although queries were resolved using few messages, messages were flooded across the network in order to update routing information every time a new node or resource was added.

It may at first not seem important to further enhance the efficiency of resource discovery, especially considering existing approaches succeed in processing requests in a number of messages already logarithmic to the number of nodes in the network. For example, reducing the messages required to process a query from an average of twelve to nine in a network of ten million nodes seems trivial.

However, when we consider the scale of P2P networks and that many users may be querying simultaneously, the seemingly small reduction appears more valid. If in the aforementioned example single queries originated on just a tenth of the network nodes, three million messages would have immediately been saved. Each message generated by the resource discovery model consumes bandwidth that could have been used to do the "real work" of the network, utilising the actual resources the query was seeking.

INTRODUCTION

P2P network overlays can be broadly divided into structured and unstructured networks. Whereas in unstructured networks nodes may join the overlay at any point, more restrictions are placed on the positions nodes can take in structured networks. A major subclass within structured networks is Distributed Hash Tables (DHTs), based on the familiar hash table data structure, which provide guaranteed lookups and typically O(log *n*) hop counts whilst maintaining only small amounts of routing state (Kelaskar, 2002). In DHTs, each node is responsible for holding one or more hash table buckets, which together form the keyspace. Particular data values (e.g. a file) can be looked up by hashing their associated key (e.g. filename) and searching the network for the node responsible for that part of the keyspace.

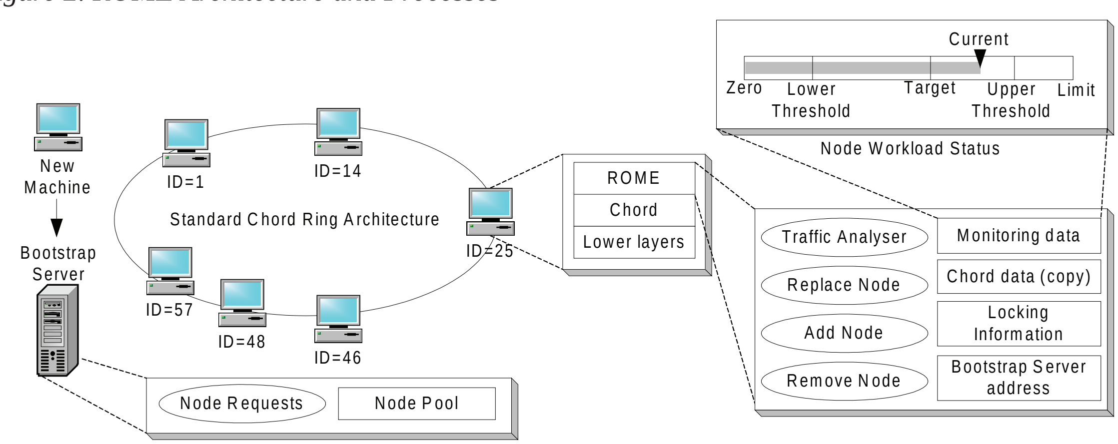
One popular DHT is Chord (Stoica, 2003), which arranges nodes into a ring overlay structure. In addition to pointers to its predecessor and successor nodes, each node maintains a small *finger table* containing routing entries for a small number of other nodes in the network. Nodes are given identifiers and become responsible for a key if their ID most closely follows the numerical hash of the key. By following successor pointers and information in the routing tables, the node holding a value associated with a given key can be found within $\log_2 n$ steps (½$\log_2 n$ on average) where n is the number of nodes in the ring. Whilst these costs are low, messaging costs should be minimised in P2P networks to allow for the maximum possible bandwidth to remain available for the "real work" of the application running on top.

In this chapter we present ROME (Reactive Overlay Monitoring and Expansion), an additional layer running above the standard Chord protocol allowing control over the construction of the network via the selection and placement of nodes within the Chord ring. ROME provides a monitoring process to detect under and overloaded nodes and add or remove nodes to service the workload. The size of a Chord ring has an impact on the number of hops it takes to resolve a lookup, with smaller rings having better performance. Our primary motivation is to provide a mechanism for controlling the size of the Chord ring to ensure it is the minimum size possible, therefore optimising the number of messages consumed. As link speeds across the internet rise but users are increasingly charged for the bandwidth they consume, we believe number of messages to be a more important metric to optimise than latency which can vary dramatically over time.

SYSTEM MODEL AND PROCESSES

Figure 2 shows an example network running ROME on top of Chord. ROME runs on each machine in the network and consists of a set of processes (ellipses) and data structures (rectangles). Processes include a traffic analyser allowing us to monitor Chord-related network traffic without needing to modify the underlying Chord protocol, and actions to add, replace and remove nodes. ROME-related data structures store monitoring data, a copy of the node's Chord data captured using the traffic analyser and ROME specific data such as the address of the bootstrap server, the node's capacity and locking information to prevent multiple ROME actions being undertaken simultaneously on a single node.

Figure 2. ROME Architecture and Processes



Additionally, we show a bootstrap server because a new machine must know the address of an existing node that it can use to establish a connection with the network. New machines contact the bootstrap server, a well-known machine that lists the addresses of the nodes currently connected to the network. This is common in P2P networks but in ROME, rather than immediately supplying the new machine with the address of the existing network node and allowing it to join instantaneously, the addresses of machines that have requested to join the network are held on the bootstrap server (node pool). Instances of ROME running on nodes already in the ring send requests to the bootstrap server for the addresses of machines to add to the ring only as they are required.

ROME ensures the underlying Chord ring is of near optimal size by monitoring the workload on each node using the traffic analyser and responding to node under or overload by adding, replacing or removing nodes. A target workload, plus lower and upper thresholds are defined for each node. In real terms, this workload is a measure of memory (to store key and routing information) and processor time/cycles (for running the protocol processes). Depending on the implementation, these capacity values are either determined by the higher-level application or explicitly defined by the user of the machine.

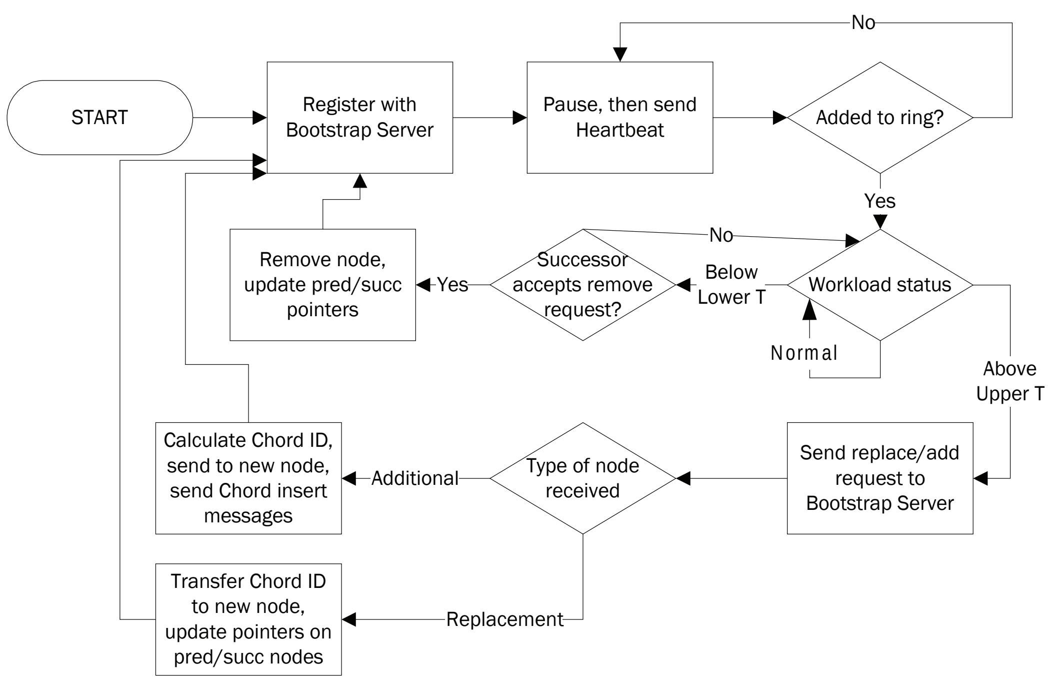
Action is taken when the current workload is either over the upper threshold or below the lower threshold, with the aim of restoring the workload as closely as possible to the target by adding, removing or replacing nodes.

Figure 3 shows a simplified flowchart of the ROME process running on a node. We discuss the various actions taken in more detail to follow.

Node Locking

Taking action has the potential to influence the workload of the node's immediate successor in the ring, so both the node and its successor must be locked before actions can be taken, thereby removing the likelihood of chain reactions and

Figure 3. Simplified ROME Flowchart



over-compensation. If action is required, ROME will send a message to the node's successor. A message will be sent back in response, indicating whether or not the lock was successful.

If the lock was unsuccessful, ROME will wait for a random exponentially increasing time period (as per the standard Exponential Back-Off Algorithm) before restarting the monitoring process. Otherwise, ROME will lock the current node and attempt to resolve the under/overload by performing an action. Placing the current node in control of the locking procedure and only seeking to lock the successor prevents potential deadlock situations from occurring.

Dependent on whether the node is under or overloaded, actions are attempted in a predefined order (detailed descriptions of each action are as follows): If a node is overloaded, ROME will send a message to the bootstrap server and attempt to replace the current node with a more capable one from the node pool. If no suitable node is available, the bootstrap server will send details of a node that can be added to the ring to take on the additional workload. If the overload has not been resolved, the process will be restarted following a wait defined by the Exponential Back-Off Algorithm.

If a node is underloaded, ROME will seek to remove the node (as long as doing so would not cause the node's successor to become overloaded). If this is not successful, ROME will wait for a time period (defined again by the Exponential Back-Off Algorithm) and restart the monitoring process.

Throughout ROME we exploit a property of Chord suggested in (Stoica, 2003), that Chord will store a node's keys on *r* nodes following the node actually responsible for that key space, so that no information is lost if the node fails. Therefore, when we modify nodes in the Chord ring, keys are automatically migrated by the underlying Chord protocol. Although ROME could migrate the keys itself, the advantage of this approach means ROME does not need direct access to any Chord data structures.

Node Joins and Network Initialisation

As would be the case for a standard Chord implementation using a server-based bootstrapping mechanism, a machine wishing to join the network contacts the bootstrap server. Our bootstrap server adds the address of the new machine in its node database, rather than providing the address of an existing network node. By preventing machines immediately joining the ring, we automatically filter out unreliable nodes that disconnect after a very short period.

The traffic analyser running on the new machine will detect the join request message, triggering ROME to send the bootstrap server a message containing the capacity the machine is prepared to set aside for the network. On machines that have registered, ROME sends periodic heartbeat messages to the bootstrap server, which allows the server to evaluate and record the machine's reliability (many missed heartbeat messages indicate an unreliable machine with low uptime).

Prior to the initialisation of an instance of the network, a bootstrap server must be defined. (We assume here the bootstrap server is external to the network itself but in implementations a machine could of course run the relevant bootstrap server processes and also be eligible to become a network node). The bootstrap server starts with an empty node database. A machine wishing to join the network then contacts the bootstrap server. The ring is immediately created with this node, as per the standard Chord process. Since no keys will have been stored at this point, the node ID cannot be determined using the methods outlined as follows. Therefore, the initial node is given node ID 0. With the network initialised, the standard processes then continue as defined.

Ring Expansion

At first glance, the obvious action to take to add capacity to the ring is to expand it by adding an additional node from the pool of available machines listed on the bootstrap server. However, it may be that there is an available machine with higher capacity that could cope with the current node's full workload. Since one of our goals is to keep the ring as small as possible, a direct node replacement is preferable (if possible), with the addition of another node to share the current node's workload as a second option.

When ROME has identified that a node is overloaded, it sends a message to the bootstrap server requesting the address of a replacement machine with a target capacity greater than its *current_workload*. The bootstrap server searches its database of available machines and, providing there is a match, replies with a message containing the address of the matching machine with the highest reliability to ROME running on the current node and removes the machine from its database. The most reliable node is found by calculating the percentage of heartbeats received from it by the bootstrap server since its initial registration:

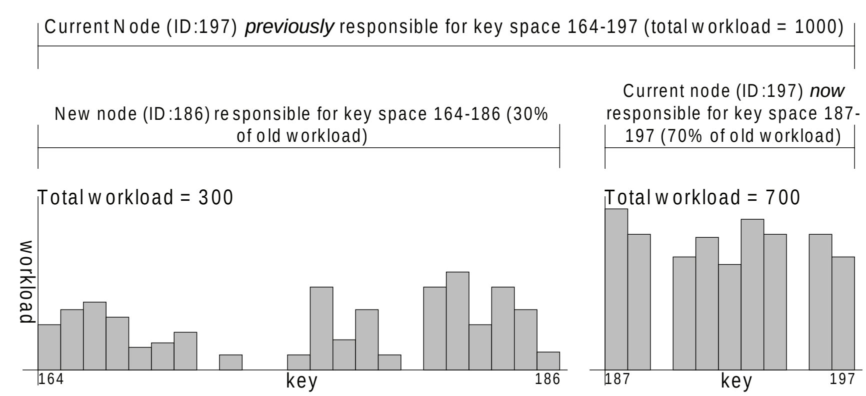
$$PercentageReceived = \frac{HeartbeatsReceived}{\frac{CurrentTime - RegistrationTime}{ExpectedIntervalBetweenHeartbeats}}$$

The node ID of the old node is transferred to the new machine by ROME running on the old node. Messages are then sent by ROME on the old node to the Chord protocol layer running on the old node's successor and predecessor nodes to alert them to update their references to the node's new IP address. Other nodes' finger table information will be updated as the standard Chord stabilisation routine is run. The migration of keys is handled by the standard Chord protocol.

If, on the other hand, the bootstrap server is unable to give the address of a machine matching the supplied capacity requirements, the second option of adding a new node to the ring to share the current node's workload must be utilised. The bootstrap server searches for a machine with a target capacity of (current_workload - current-node_target_capacity). Once the overloaded node has received a message containing the address of a machine to be added to the ring, it must organise for its keys to be shared between the current node and the new node. This is achieved by inserting the new node into the ring between the current node and its predecessor, giving the new node a node ID such that Predecessor_ID < New_Node_ID < Current_Node_ID.

The actual node ID is chosen based on the cumulative workload generated by each key on the current node and how much workload must be moved to the new node. Figure 4 shows how 300 units of workload could be moved to a new node from the current node by (in the example) transferring keys 164-186 to the new node. Similar to the aforementioned, a message informs the new node of its node ID. The new node is inserted into the ring by ROME on the current node sending insert

Figure 4. Example key transfer from current node to new node during node addition



messages to the standard Chord protocol running on itself, its successor and its old predecessor. Again, key migration is handled by the Chord protocol.

Node Removal

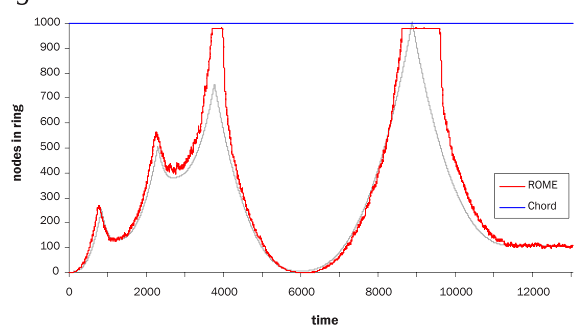
The actions introduced are mechanisms to cope with nodes becoming overloaded and to better organise workload within the ring. However, it is also possible that workload may decrease to an extent that the underload cannot be rectified through re-organisation, making it unnecessary to keep all nodes in the ring.

When ROME wishes to remove its host node, it sends a remove request message to its successor containing its current workload. ROME on the successor checks whether the node's workload combined with the successor's current workload is over the successor's upper threshold. It responds with a message indicating whether the remove operation can take place. If the node can be removed, ROME sends a message to the Chord protocol on its predecessor informing it its successor is now the current node's successor. Similarly, it sends a message to its successor informing it its predecessor is now the current node's predecessor. Finally, ROME sends a message to the bootstrap server, which triggers the node to be re-added to the pool of available nodes.

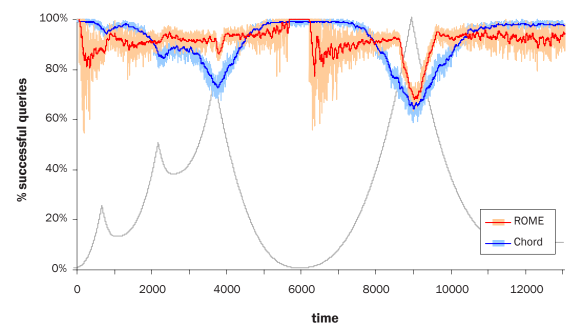
EVALUATION

To evaluate ROME we built a simulator in Perl. The simulator supports two Chord rings, one emulating a standard implementation of Chord and one emulating a Chord ring with ROME running on top. This allows a direct comparison of ROME with standard Chord and the benefits or increased costs to be explored. 1000 nodes were made available to each ring. The ROME-specific

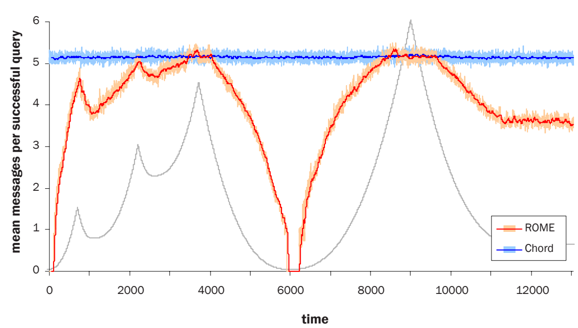
Figure 5. Simulation results



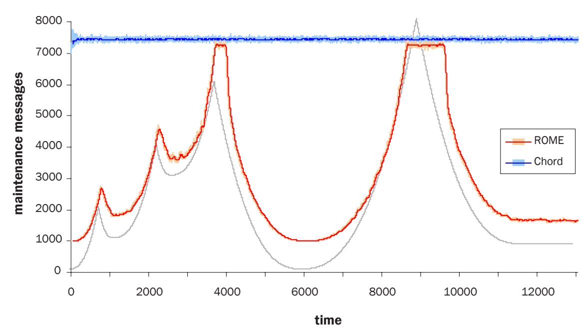
a) Variation in Number of Nodes in Each Ring



b) Percentage of Queries Successfully Resolved



c) Mean Number of Messages per Successful Query



d) Maintenance Cost

node thresholds were set as follows: A node was considered underloaded if its workload fell below 5% of its total capacity, with the upper threshold has been set at 95%. The target workload was set at 50% node capacity. 500 lookups were randomly generated and executed each clock tick. The experiment was repeated 3 times to ensure no anomalies were present. During the experiment the network workload was varied over time as indicated by the grey line superimposed on the graphs in Figure 5, up to the total capacity of all available nodes. This distribution was decided upon because it allows the experiment to show how the ROME-enabled network behaves under different workload conditions.

One of the key functions of ROME is its management of the size of the underlying Chord ring. The potential message savings are directly derived from this. The number of nodes in the network running Chord (blue) and the network running ROME (red) at every clock tick throughout the experiment is shown in Figure 5a.

As expected, the number of nodes in the Chord network remains constant throughout since no management of the volume of nodes is included in the standard Chord protocol. In contrast, using ROME the number of nodes varies with the change in network workload.

The number of nodes in the ROME-managed ring never reaches 1000 because of the effects of node churn. Whereas nodes are immediately added to the Chord ring when they join, with a ROME-managed network these nodes are first added to the node pool. Therefore, with 10 simulated node failures and 10 node joins per clock tick, at most approximately 980 nodes will be present in the ROME ring with another 20 remaining in the node pool.

In the ROME-enabled network, all nodes are added to the ring (with the exception as described previously) when the network-wide workload approaches 75% of the total capacity of all the available nodes. A small fluctuation in the number of nodes can be seen even when the workload is kept constant at the end of the experiment (from approximately clock tick 11500). This again is due to node churn: Failures cause some nodes to disappear from the ring and sometimes better nodes are available in the pool to provide replacements for these due to higher capacity nodes joining the network. This results in the workload of two or more failed nodes being supported by a single replacement. The opposite also applies, being that sometimes extra nodes must be added to the ring due to a single replacement node not being available.

Figure 5b shows the percentage of queries that were successfully completed each clock tick in the Chord and ROME-enabled rings. The light blue and tan lines show the actual percentage recorded at each iteration, while the darker blue and red lines plot a 50 point moving average of the data. We define a successful *query* as a lookup successfully routed that can also be answered by the destination node (overloaded nodes are forced to drop queries they cannot handle, so even though a query reaches the destination it may not be successfully processed).

Due to ROME's management of node overload, the ROME-enabled ring exhibits a lower percentage of failed queries than Chord for the majority of the experiment, except where only a small number of nodes exist in the ring. However, since these failures occur when the ring is small and thus the number of messages taken to process them is also small (see Figure 4c), it could be argued that they could be re-run without incurring an exceptional cost.

The major goal of ROME is to reduce the number messages required to process queries in Chord. As shown in Figure 5c, this is largely achieved although there are some minor fluctuations above Chord around time 4000 (when workload is 75%) and time 8500 (again, when workload is over 75%). It is reasonable to assume that network owners may wish to consider whether running ROME is appropriate if they expect their network's workload to consistently approach ring capacity.

An average of 591.4 messages were saved using ROME each time tick based on the 500 queries submitted (with -190 the minimum and 2663 the maximum savings).

In order for ROME-enabled rings to truly provide message savings over standard Chord rings, the lookup message savings as shown before must not be at the expense of increased maintenance message cost. Maintenance messages include those generated by the standard Chord join and stabilisation routines, plus ROME-specific messages used to transmit the necessary data and instructions for actions to be performed and heartbeat messages sent to the bootstrap server from nodes currently residing outside the ring. The maintenance costs of the ROME-enabled ring and the equivalent Chord ring are shown in Figure 5d. The total cost of ROME actions plus maintaining the underlying Chord ring is consistently less than the maintenance cost of standard Chord. The cost includes heartbeat messages (one every clock tick for every node in the bootstrap server's node pool), meaning ROME's minimum maintenance cost lies at approximately 1000 messages rather than at zero as might initially be expected. On average, 4423.4 messages were saved per time tick with a minimum saving of –35 messages and a maximum of 6746.

CONCLUSION AND FUTURE WORK

We have shown that, by adding a layer to control node joins and processes to monitor node workload, we are able to construct smaller than standard Chord rings with the benefit of smaller messaging costs throughout the network. Although we add more functionality to the bootstrap server, this is only utilised for node join, replacement and removal functions. As with a standard Chord implementation, the ring is still capable of performing its core functions if the server becomes unavailable. Through simulation we have proved that construction of small rings is desirable, and demonstrated this is possible using ROME when the total capacity of machines available to become members of the ring exceeds the workload placed on the ring.

The results gathered show that ROME can be best applied in situations where the sum of the capacities of all the nodes wishing to connect to the network is much greater than the workload being placed on the network. If these are near the same value, most or all the nodes must be added to the underlying Chord ring and therefore ROME provides little benefit. In some cases where the workload and capacity are near identical, the increased number of changes to the ring caused by ROME's actions can lead to the message cost of lookups exceeding that of a standard Chord ring.

Application developers should consider adoption of ROME in applications where workload is likely to be lower than the capacity offered by the connected nodes. An obvious example of this is a simple distributed lookup application: A user would need to be connected to the network to issue a lookup, thus there are likely to be many machines wishing to join the ring (although these can be hosted in the node pool rather than be members of the underlying Chord ring). The workload associated with each lookup is likely to be low so the total network-wide workload could be handled by a small number of machines. Utilising ROME for this application would enable lookups to be routed in fewer messages than in a standard Chord ring where each machine would have to be a member of the ring.

A widely held assumption has been that nodes wishing to participate in a Peer-to-Peer network immediately join the network. Through this work we have shown that it is possible to reduce the message cost associated with a network by controlling the addition and removal of nodes to ensure the overlay contains only enough nodes to service the current workload. This has implications for the design of future Peer-to-Peer network overlays and protocols.

ROME could be used as the foundation for building new classes of application such as real-time searching technologies to allow peers to share any type of resources or complex collaborative applications which take the notion of collaboration further than the existing fairly simplistic interactive whiteboard applications (MacDonald, 2003). Although ROME has been presented as an optimisation within Peer-to-Peer networks, it may also be applicable in other domains such as wireless or ad-hoc networks which have a similar property of dynamic machine joins and leaves as in P2P. Especially where networks are made up of low bandwidth connections, the number of hops between points in the network needs to be minimised. Maximising usage of devices within such environments would ensure the network (and the investment behind it) is fully utilised.

ROME could also be linked in with a mechanism supporting High Throughput Computing on distributed computing resources. One such mechanism is Condor (Litzkow, 1998), which sends a user-submitted job for execution on a remote machine within a pool of available machines. This appears similar to the concept of the node pool on a Bootstrap Server. ROME operations could be developed to run on the remote machine with a possible view to utilising add, replace and remove actions to manage the job workload.

The underlying Peer-to-Peer network protocol used within this work has been Chord. However, it is likely that the method of controlled construction of the network overlay can lead to message savings in other structured DHT-based Peer-to-Peer networks where the lookup cost is proportional to the number of nodes in the overlay. Calculations related to placement of new nodes to ensure they take on the correct proportion of workload will be more difficult in DHT-based overlays which rely on multi-dimensional keyspaces rather than the one-dimensional keyspace found in Chord, but this is merely a more complex problem to solve rather than something insurmountable.

It may also be possible to expand the ideas presented in ROME to apply in the domain of unstructured P2P networks. However, there are many different issues to be addressed before this could be attempted. In unstructured networks there is no concept of keyspace and nodes dealing with workload or hosting files relating to their area of keyspace. Another form of organisation would need to be implemented or adapted to allow for workload to be moved around in these networks so that nodes could take on extra workload and under-utilised nodes could be removed from the overlay.

Peer-to-Peer networks are in a collection of technologies that are shifting power towards users in what can be viewed as the democratisation of the Internet. They bring new challenges to stakeholders, whether they are ISPs, network administrators, developers or end users. However, they also bring new opportunities to produce applications not possible in the traditional client/server based Internet environment. In the future we may see P2P combined with emerging technologies such as wikis, podcasts, semantic Web, application service providers and others, allowing for evermore exciting advances. However, all these applications will rely on the underlying network overlay at the application layer. The work in this chapter presents a possible extension to this base service, allowing more efficient lookups to take place in structured DHT-based networks. Other topologies will doubtless be developed in the future, but the idea of controlled construction of the network overlay leading to reduced messaging costs where these are proportional to network size will remain.

RELATED WORK

Many extensions and modifications to the standard Chord protocol have been proposed, several of which share our aim of reducing the number of hops required to process a lookup. However, to our knowledge none of these do this by controlling the expansion of the Chord ring itself. In fact, ROME is complementary to these extensions. Using them together may provide even greater message savings.

Kaashoek and Karger (2003) combine Chord with de Bruijn graphs to produce Koorde, allowing for varying amounts of routing information to be stored and thus achieving hop counts between O(log *n*) and O((log *n*)/log log *n*), where *n* is the number of nodes in the network. It is clear that a reduction in *n* would also reduce Koorde hop counts. The authors also suggest that finding a system that is load balanced and degree optimal (optimising the amount of routing information stored) is an open question. With our control over node joins, small ring size and placement of nodes, we believe ROME can go at least some way to achieving this.

Single hop lookups are achieved by storing a global routing table on every node by Leong and Li (2004), but only in circumstances where network churn is low, something rarely achievable in realistic P2P implementations.

Other researchers have explored creating hierarchies of multiple Chord rings. Garcés-Erice et al (2003) demonstrate reduction in hop counts through their assumption that high quality nodes are always available to add to the top layer of their hierarchy. Mislove and Druschel (2004) build sub-rings within administrative domains, exploiting node locality to reduce latency of lookups. Locality could also be exploited using our protocol in the single layer rings, becoming an extra selection criteria when searching the node pool for machines to join the ring.

Ratnasamy et al (2002) reviewed several DHT-based routing algorithms and direct researchers toward several open questions. They suggest that node heterogeneity is extreme, which is a property we use to our advantage. By allowing only the best nodes to join the ring, we build a ring with less probability of failure than one including all available nodes.

REFERENCES

Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., & Werthimer, D. (2002). SETI@home: An Experiment in Public-Resource Computing. *Communications of the ACM*, *45*(11), 56-61.

BBC News. (2005). Legal fight hits music pirates. *British Broadcasting Corporation*. Retrieved November 14, 2005, from http://news.bbc.co.uk/1/hi/entertainment/music/4438324.stm

Bohannon, J. (2005). Grassroots Supercomputing. *Science*, *308*(5723), 810-813.

Digital Spy. (2005). Report: Illegal file sharing on increase. *Digital Spy Limited*. Retrieved November 16, 2005, from http://www.digitalspy.co.uk/article/ds26272.html

Garcés-Erice, L., Biersack, E. W., Ross, K. W., Felber, P. A., & Urvoy-Keller, G. (2003). Hierarchical peer-to-peer systems. *Parallel Processing Letters*, *13*(4), 643-657.

Kaashoek, M. F., & Karger, D. R. (2003). Koorde: A simple degree-optimal distributed hash table. In *Proceedings of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS03)*. Berkeley, CA.

Kelaskar, M., Matossian, V., Mehra, P., Paul, D., & Parashar, M. (2002). A study of discovery mechanisms for peer-to-peer applications. In *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid Workshop on Global and Peer-to-Peer on Large Scale Distributed Systems* (pp. 444-445). Berlin, Germany.

Leong, B., & Li, J. (2004). Achieving one-hop DHT lookup and strong stabilisation by passing tokens. In *Proceedings of the 12th International Conference on Networks (ICON2004)*. Singapore.

Litzkow, M., Livny, M., & Mutka, M. (1998). Condor – A Hunter of Idle Workstations. In *Proceedings of the 8th International Conference* *of Distributed Computing Systems (ICDCS 1998)* (pp. 104-111). San Jose, CA, USA.

MacDonald, M. (2003). *Peer-to-Peer with VB .NET*. New York: Springer-Verlag.

Minar, N., & Hedlund, M. (2001). A Network of Peers: Peer-to-Peer Models Through the History of the Internet. In A. Oram (Ed.), *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology* (pp. 3-20) Sebastopol, CA, USA. O'Reilly.

Mislove, A., & Druschel, P. (2004). Providing administrative control and autonomy in structured peer-to-peer overlays. In *Proceedings of the 3rd International Workshop on Peer-to-Peer Systems (IPTPS04)*. San Diego, CA.

Motion Picture Association. Belgian & Swiss Authorities Break Razorback2 [press release]. MPA, 21 February 2006.

Ratnasamy, S., Shenker, S., & Stoica, I. (2002). Routing algorithms for DHTs: some open questions. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS02)*. Cambridge, MA.

Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., & Balakrishnan, H. (2003). Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on Networking*, *11*(1), 17-32.

Zhao, B., Huang, L., Stribling, J., Rhea, S., Joseph, A., & Kubiatowizc, J. (2003). Tapestry: A resilient global scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, *22*(1), 41-53.

KEY TERMS

Distributed Hash Table: A method for storing hash tables in geographically distributed locations in order to provide a failsafe lookup mechanism for distributed computing.

Fault Tolerance: The ability of a system to respond gracefully to an unexpected hardware or software failure.

Hash Table: A lookup table that is designed to efficiently store non-contiguous keys that may have wide gaps in their alphabetic and numeric sequences.

Network Overlay: An overlay network is a computer network which is built on top of another network. Nodes in the overlay can be thought of as being connected by virtual or logical links, each of which corresponds to a path, perhaps through many physical links, in the underlying network.

Network Traffic: Data transmitted over the network at a given moment.

P2P: A distributed network architecture may be called a Peer-to-Peer network, if the participants share a part of their own hardware resources (processing power, storage capacity, network link capacity, printers). They are accessible by other peers directly, without passing intermediary entities. The participants of such a network are thus resource providers as well as resource requestors (Servent-concept).

Query Hops: The number of times the query has been forwarded.

Chapter LVII
Designing a New Service Overlay on a Carrier Network Using the Efficient Frontier

Susan J. Chinburg
Rogers State University, USA

George Scheets
Oklahoma State University, USA

ABSTRACT

The network carrier must utilize the full potential of existing physical infrastructure before any new investments can be considered. Much work has been done to develop optimal design methodologies that generate a "best" design but designs often have to be changed to meet operational considerations, potentially mitigating any optimization benefits. This work proposes to expand the optimal network design process by applying a process borrowed from Finance and Operations Research/Management Science literature known as the Efficient Frontier or production frontier analysis. Using this idea, a portfolio of designs differing in cost, and number, size and location of nodes will be developed and from that the efficient frontier will be defined. By comparing any subsequent design to the frontier, the network designer can have an understanding of the impact of changes in the design strategy to the long-term cost effectiveness.

I. INTRODUCTION

Mitigating the risk of expanding a network in today's economic climate requires that the network carrier utilize the full potential of existing infrastructure before any new investments are considered. The design of actual networks is the product of multidimensional constructs of performance and reliability optimized around a certain factor or factors such as cost or reliability. Typically, a methodology is created to define near optimal solution depending upon the constraint parameters to be considered. Much work has been done to develop optimal design methodologies using one or multiple objectives but these problems are intractable and NP-hard and can require complex computational solutions that are often best approximations (Yun et al, 2001). Even after the optimal design has been developed, during implementation the design often has to be changed to meet operational considerations, potentially mitigating any optimization benefits. Design of telecommunications networks is a complex, usually iterative, process with several objective functions such as low cost, reliability, security, and survivability. To find the optimized design, an exhaustive analysis must examine every possible combination of all parameters thereby proving the best fit to meet the specified design characteristics. Finding near-optimal solutions requires fast heuristic algorithms, establishing acceptable constraints or bounds and then "relaxing" these constraints so the solutions can be found with reasonable resources and time (Kennington et al 2007). The bulk of published research and literature in mathematical modeling and optimization deals with ways of making this very complex problem simpler and easier to solve. Mathematical modeling and optimization are well-developed and mature areas of research with a variety of articles available for the interested reader who is referred to Bertsekas (1998), Sanso and Soriano (1999), Grover and Doucette (2001), and Kennington et al (2007) for broader overviews.

This paper is intended to discuss the impact of applying the efficient frontier concept to network design of a new service overlay and assumes that a method for defining the optimal network design has been adopted. The method used to calculate the optimal design and cost of the case studies in this chapter will be presented, but that method is not the main focus of this effort. Applying the concept of the efficient frontier is the main focus of this chapter.

Using the efficient frontier or production frontier, a concept borrowed from finance, agriculture economics, and operations research literature, a design process can be established that creates not just a single optimal solutions but also a set of solutions, (Markowitz, 1959 and Farrell, 1957). A brief discussion of the Efficient Frontier concept will be presented but interested readers are referred to Fare et al (1994) and Copeland et al. (2005) for mosolutions for the problem. By developing a set of optimal solutions using design heuristics that best fit the given project, an envelope of cost functions is created. The lower boundary of this envelope is the efficient frontier or suite of "best" designs as far as the parameters used for the optimization. The efficiency of any design is the relationship of the final cost of a design to the frontier. Evaluating the distance any point is from the efficient frontier gives a measure of how efficient the design is as related to the parameters used in the optimization. If the cost of a design lies upon the frontier it is considered efficient, the further off the frontier, the less efficient the design. By understanding the impact of design changes to the final cost of the network, well-informed design changes can be implemented with regard to final or long-term cost.

Complicating the design process for communication networks is that there are often multiple objectives such as reliability, security, and recoverability in addition to low cost. For multiple-objective optimization problems a set of Pareto-optimal solutions can usually be defined (Sawaragi et al 1985 and Koopmans 1951). The set of Pareto solutions becomes the efficient frontier for the design process. Pareto optimal solutions do not necessarily optimize all objective functions but create a set of candidate solutions that can be individually evaluated for use. The input parameters to these problems are usually arrived at through analysis related to the subject area being investigated (Yun et al 2001). The method presented by Yun et al (2001) involves extensive mathematical multiple objective optimization utilizing much computational resources. This paper presents a similar process in that an efficient frontier is created using a set of optimal single objective optimization solutions developed from sensitivity analyses.

II. Study Methodology

In this section an outline of a set of experiments conducted with a mixed integer linear programming (LP) problem formulation of the service overlay with growth problem (SOGP) is presented. The SOGP was formulated to solve the problem; Overlay System Cost = Overlay Switch Cost + Overlay Link Cost, and provides a general formula useful for examining the installation of a new network overlaid on top of an already existing network. This problem is expressed by the following equation.

Overlay System Cost

$$= \sum_{i=1}^{n} [x_i (C_1 + K_1 * (node_i\_traffic)^{\alpha})] \quad \text{[switch chassis costs]}$$

$$+ \quad K_2 \sum_{i=1}^{n} x_i (node_i\_traffic)^{\beta} \quad \text{[traffic or connections costs]}$$

$$+ \quad K_L * \sum_{i=1}^{n} \sum_{j=1}^{n} (traffic_{ij})^{\chi} * (distance_{ij})^{\delta} \quad \text{[link costs]} \quad 1$$

$_i$ = indicates which node from the set N = {1, 2,…,n} where n is the total number of nodes in the overlay design.

$i,j \in N = \{1, \ldots n\}$. x_i = binary constant, (0 or 1) indicating whether or not an overlay switch is located at node i.

C_1 = initial cost of chassis

K_1 = arbitrary unit cost of a chassis $node_i\_traffic$ = the amount of traffic flow that will move through the node i. Also called the number of connections needed at a switch.

K_2 = an arbitrary unit cost assigned to each connection.

K_L = an arbitrary unit cost factor.

$traffic_{ij}$ = link capacity from node i to node j.

$Distance_{ij}$ = route miles from node i to node j.

α = Exponential factor that modifies the impact of traffic flow upon the size of the switch chassis. $\alpha < 1$.

β = an exponential factor that will describe the shape of the curve, $\beta < 1$. χ and δ; exponential factor for traffic and distance respectively. For this study χ and $\delta = 1$ but other cost models could have these values less than one. For instance, some Frame Relay service providers do not price connectivity in terms of distance, so δ could be equal to zero. Other cost models with $\chi < 1$ could show economies of scale related to capacity. The closer α is to 1, the closer the function is to linear.

Various nodal distribution strategies for an overlay were examined by manually controlling the node input to the design for the four case study legacy networks. The basic assumption in this investigation was that an underlying legacy network topology (nodes and connections between nodes) exists and a new service overlay was to be built on top of the existing network structure with no additional physical links or nodes added. New service traffic flows were transmitted over the existing topology.

SOGP, modeled after Grover and Doucette 2001 and Chang and Gavish 1993, creates a sum of shortest paths for the traffic flow solution for the input design criteria of nodes, links and traffic flow values.

Definitions:

- n is the number of nodes, **N** is the set of such nodes. **N** = {1, 2, 3, … , n} $i,j \in N$.
- **A** is the set of possible bi-directional edges in the graph on the set of nodes **N**.
- **T** is the set of time periods, t, in the study. **T** = {1, 2, … , t}.
- $\mathbf{D}^t$ is the set of all non-zero flow demand quantities exchanged between nodes, in-

dexed by *r*. D will be different for each time period t. The elements of $\mathbf{D}^t$ represent the traffic to be carried by the network for each city pair.

- d^{rt} is the amount of flow demand associated with the *r*th demand pair in **D** for period, t.
- O[*r*] is the node that is the origin for the *r*th flow demand pair in **D**.
- TR[r] is the corresponding target or destination.
- cl_{ij} (= cl_{ji}) is the incremental cost of adding one unit of capacity to edge (i,j). These incremental steps can be the same or different for each time period.
- w^{rt}_{ij} is the amount of working flow of the rth demand routed between nodes (i,j) on link i,j or j,i for period t.
- lw_{ij}^t is the working capacity assigned to the edge between nodes (i,j) to support all working flows routed over that edge for period t.
- S_j^t is the fixed cost of a switch at node j.
- S_{jk}^t is the fixed cost of a switch at node j for size increment k. There are incremental steps, k, in this function determined by the number of ports needed to carry all the traffic flow coming through and into the switch at node j. These incremental steps can be the same or different for each time period. The value of S_{jk} at each increment of k will be arbitrarily defined.
- cp_j is the incremental cost of adding one unit of capacity to switch at node j. These incremental costs can be the same or different for each time period.
- P_j^t = the number of connections needed in switch at node j to accommodate all the traffic flow that will enter or leave this switch. It is the sum of the demand flow routed to/from and through node j.
- CS_{jk} is the incremental cost of the switch chassis. CS_{jk} is a constant value set before the model is analyzed. These incremental costs can be the same or different for each time period. This formulation has them the same for each node.
- K is the set of incremental costs, k corresponding to the number of different sizes of chassis used in the model, b.. **K** = {0, 1, 2, …, k}.
- CN_{jk} is the maximum number of connections that any switch size will allow. For each k there will be a predefined constant such as 1000, depending on the size of the switch chassis.
- M is an arbitrarily large constant that is greater than any capacity needed to accommodate all the flow through a switch.
- z_{jk} is a binary variable (0,1) used to facilitate the step cost function of the switch chassis. The number of different sizes of the chassis, b, is the number of different values for the subscript k. K ={0,1,2, …, k). z_{jk} will be 1 for the appropriate size of P_j and zero for the rest. Only one chassis will be installed at each node. When k = 0 there are no (0) connections for that nodes and therefore no switch at that nodes.
- x_j is a binary variable (0,1) used to control the nodes that are allowed to have a switch in the design. This variable can be manually controlled.

Objective Statement: (see Box 1.)

The objective statement 2 seeks to minimize the cost of the network overlay based on three factors, the cost of the node/switch chassis, the cost of each connection to the switch chassis and the cost of additional capacity on a link to support the required flow. Equations 3, 4, and 5 are standard flow balance constraints that require that all the flow leaves the source, all the flow arrives at the destination and that at all transshipment nodes the flow is balanced so that all the flow that enters the nodes leaves the node. Equation 6, the link capacity constraint, sums all work-

Box 1.

Minimize

$$\sum_{t\in T} \left\{ \sum_{j\in N} x_j S_j^t + \sum_{j\in N} x_j \{cp_j^t * P_j^t\} + \sum_{i,j\in A} \{cl_{ij}^t * lw_{ij}^t\} \right\} \quad 2$$

Subject to:

$$\sum_{j\in A} w^{rt}_{ij} - \sum_{j\in A} w^{rt}_{ji} = d^{rt} \text{ for all } r \in \mathbf{D}, i = O[r] \text{ and for all } t\in T. \quad 3$$

$$\sum_{\in A\, j} w^{rt}_{ij} - \sum_{\in A} w^{rt}_{ji} = - d^{rt} \text{ for all } r \in \mathbf{D}, i = TR[r] \text{ and for all } t\in T. \quad 4$$

$$\sum_{\in A} w^{rt}_{ij} - \sum_{j\in A} w^{rt}_{ji} = 0 \text{ for all } r \in \mathbf{D}, \text{ for all } i \notin \{O[r] \text{ or } TR[r]\} \text{ and for all } t\in T. \quad 5$$

$$lw_{ij}^t = \sum_{i,j\in A} w^{rt}_{ij} + \sum_{j,i\in A} w^{rt}_{ji} \text{ for all } i,j\in \mathbf{A} \text{ and for all } t \in \mathbf{T}. \quad 6$$

$$P_j^t = lw_{ij}^t + lw_{ji}^t \text{ for all } i,j\in \mathbf{A} \text{ and for all } t \in \mathbf{T}. \quad 7$$

$$P_j^t <= M * (z_{j0} + z_{j1} + z_{j2} + \ldots + z_{jk}) \quad 8$$

$$z_{jk}^t \in \{0,1\} \text{ for all } j\in\mathbf{N}, \text{ for all } i,j \in \mathbf{A} \text{ and for all } t\in T . \quad 9$$

$$P_j^t <= 0z_{j0} + CN_{j1}\, z_{j1} + CN_{j2}\, z_{j2} + \ldots + CN_{jk}\, z_{jk} \quad 10$$

$$S_j^t = 0z_{j0} * CS_{j0}^t + z_{j1}^t * CS_{j1}^t + \ldots + z_{jk}^t * CS_{jk}^t \quad 11$$

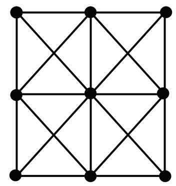
$$\sum z_{jk}^t = 1, \text{ for all } j \text{ and for all } k \text{ and for all } t \in \mathbf{T}. \quad 12$$

$$lw_{ij}^t - lw_{ij}^{t-1} >= 0, \text{ for all } i,j \in \mathbf{A} \text{ and for all } t \in \mathbf{T}. \quad 13$$

$$w^{rt}_{i}, w^{rt}_{i} >= 0. \quad 14$$

ing flow, $w^{rt}_{ij,}$ and $w^{rt}_{ji,}$ that traverse this link to determine the amount of capacity needed on the link. Equation 7 is the switch size constraint that sums all working flows that go though this link to determine the number of connections needed for the switch at node j. Equation 8 says that the value of P_j^t will be either zero if there is not a switch at this location or it will be less than M, that is a number sufficiently larger than any connection capacity needed at any switch. Equation 9 defines the binary choice variable, z_{jk}^t, defined for each node, j, as to which size of switch chassis is installed at the node. Equation 10 determines the correct size of switch to be placed at a node based on the number of connections required to deliver the traffic flow. Equation 11 calculates the cost of the switch based on the previous constraint. Since this is a minimization problem the cheapest size will be selected based on the correct value of z_{jk} from constraint 9. Equation 12 requires that there be only one switch size for each node if any at all. Equation 13 requires that the input to the next period, t, be the output of the previous period. And equation 14 states that the

Figure 1. Initial 9 node proof of concept model



working flow, $w^{rt}_{ij,}$ and w^{rt}_{ji}, will be either zero or a positive number.

II.2 Case Study Topologies

The initial legacy topology for proof of concept testing was a 9-node model (Figure 1).

Three case studies then were conducted using a generic North American network based on that of WilTel (Figure 2), the NSFNet from the US research network, the beginnings of the Internet (Figure 3), and a generic Pan-European model based on that used in Maesschalck et. al., 2003 (Figure 4). The North American network has 27 nodes and 43 links. The Pan-European network has 28 nodes and 43 links. The NSFNet network model has 15 nodes and 22 links.

The traffic matrix was developed using the population-based function as described by Cahn (1998). Population data from the US Census database was used for the North American and the NSFNet case studies. For the Pan-European topology the populations of the metropolitan city

Figure 2. The North American legacy topology based on the WilTel North American network. The number by each city name is an index

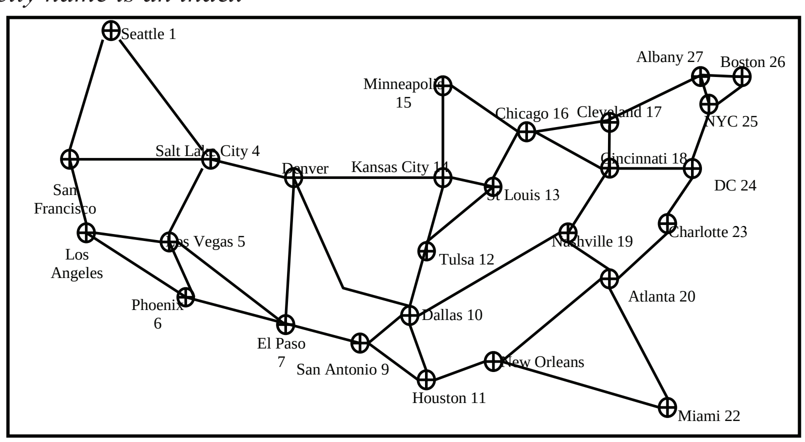


Figure 3 The NSFNet legacy topology. The number by each city is an index

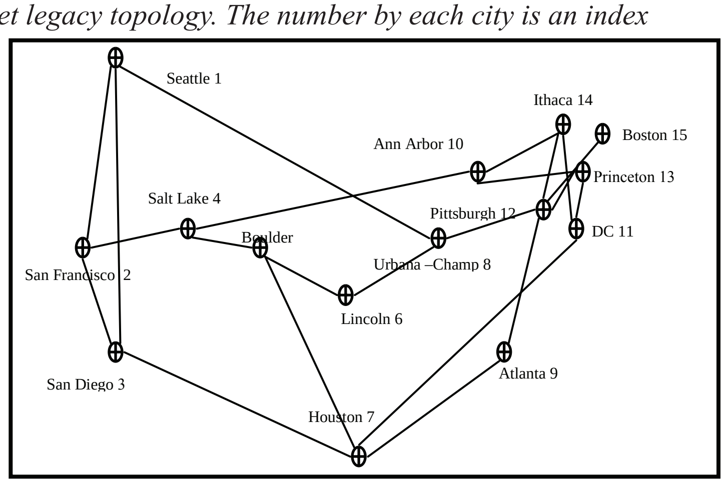
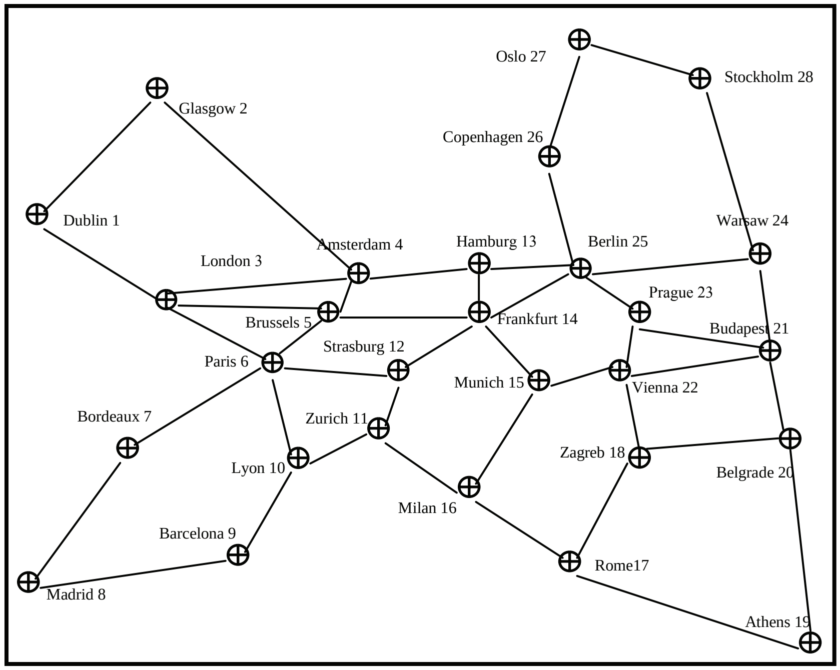


Figure 4. Pan-European legacy topology. The number by each city is an index



areas were from data supplied by European city mayors (http://www.citymayors.com/features/euro_cities.html). Distance between each city pair was calculated using the Microsoft VirtualGlobe, 1998 edition.

Three cost functions were used in this study, the cost of capacity on a link, the cost of the switch chassis and the cost of each flow connection in and out of the switch chassis. Cost models for the overlay were developed using generic cost functions based on previously published data from WilTel and other publicly available information. These functions were built from the network carrier perspective, essentially, the cost required to implement the new service overlay. Each unit of connection in the switch is a constant cost which in this study was arbitrarily set to 125. The actual switch chassis costs are a step function that is not generally linear and were set to \$1000, \$2000, \$4000, \$10000, and \$20000 with switch sizes set to 1000 ports, 2000 ports, 4000 ports, 12000 ports, and 24000 ports respectively. Each unit of link cost is a function of the distance of the link and the link capacity times a dollar amount. The distance was determined by summing the linear distances between the nodes traversed by the overlay link. For this study to keep link costs simple, the dollar amount was arbitrarily set to \$1 per distance unit. Capacity multiplied by distance and by the unit link cost created the total link cost for each link.

The SOGP formulation was implemented in single sequential periods with the topology output, including number, location and size of switches, the links and the capacity of links of one growth period as the input for the next growth period. The only difference between each period was the increase in the amount of traffic. Each node in the legacy topology would contribute traffic to the design, in the SOGP implementation; chassis and ports were allowed at all nodes including all the access-only nodes. These access-only node costs, chassis and ports, were manually deducted from the final costs to determine the final total cost for the overlay design.

The number and distribution of switches in the nodal distribution pattern in the service overlay were manually varied using different design strategies developed around both number of switches, node population size and location of the nodes in the backbone. To approximate network growth a five level traffic matrix was developed. 1X growth stage represented the traffic matrix calculated as described in Step 2 of this section. Four additional growth stages were calculated by multiplying the initial amount of traffic demand, 1X growth stage, by 1.5, 3.5 and 10 creating five growth stages 1X, 1.5X, 3X, 5X, and 10X. For this study each design strategy was modeled using all five-growth stages.

The location of the backbone design for each scenario was established using the shortest path connections between each backbone node. Variations in design strategies were based on well-accepted design heuristics that would be common implemented, including number of overlay switches, edge versus central locations and population of the selected node (Cahn 1998). To control the population variable three groups of nodes were established based on population size of the city at the node, small, medium, and large, for each real world case study.

III. Results of Case Studies

Four separate case studies were evaluated, first, a small proof-of-concept 9-node legacy network (Figure 1) and then three larger closer-to-real-world legacy models, a North American, NA, model (Figure 2), the NSFNet (Figure 3), and a Pan-European, PE, model (Figure 4). This section presents the results of the experimental analyses. The primary network carrier nodal backbone design strategy used to develop test network configurations used the philosophy that once a legacy node was in the overlay backbone the overlay switch handled all traffic flows passing through this node. Hereafter this will be referred to as design strategy 1. A secondary strategy 2 created a mesh topology with logical links between each node pair in the overlay backbone. The only traffic flow handled by an overlay switch in the full logical link mesh is traffic flow that originates and terminates at that switch location. Those nodes not in the overlay backbone will have access to the nearest overlay backbone switch and that switch will handle (originate and terminate) their traffic flow. Thru traffic flow passing through a node with an overlay switch will be processed by the assets of the underlying legacy network in design strategy 2. A limited number of mixed topologies, using a combination of legacy and logical links, were also evaluated.

III.2 The 9-node Case Study

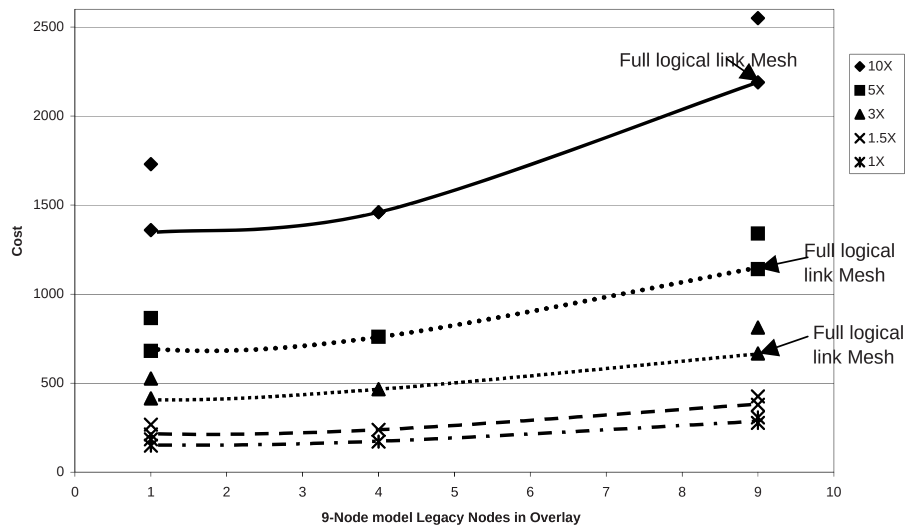
The 9-node network topology (Figure 1) is well connected with an average degree of connectivity of 4.4. Link distances for this model are an arbitrary value of 1 or 2 with an average of 1.4 and the switch costs are an arbitrary 10, 20, 30, 40 and 50 for sizes of 50, 100, 200, and 500 connections respectively. The unit cost constants of switch costs, K_S, are much greater than link costs, K_L, $K_S >> K_L$.

The most cost effective overlay design evaluated for this model is the one switch centralized in a star pattern (Figure 5). An intermediate distribution of nearly 45% (4 out of 9) of the legacy

Table 1. Network characteristics of case study legacy networks. Network characteristics of number of nodes, connectivity of nodes, population of nodes, number of links, link length are presented in this table. C is the connectivity of the nodes in the network

Network	Nodes	C	Pop	Links	Link length
	#	Average	Ave in M	#	(Mean)
9 node	9	4.44	NA	20	1.40
PE	28	3.04	1.42	43	248.21
NA	27	2.93	3.79	43	436.95
NSFNet	15	3.19	3.92	22	1085.45

Figure 5. Experimental total cost results for the 9-node legacy network. One switch on a central location was the lowest cost solution evaluated by this study while one node switch located at a periphery location was more costly



nodes in the service overlay is the next most cost effective long-term approach evaluated.

For the 9-node legacy topology, overlay designs with switches at all nodes were evaluated using both design strategies. There were switches at all nodes with logical links between all nodes, design strategy 2, and the other design used the links of the legacy topology, design strategy 1. The centralized location will have minimized link costs because all traffic flow paths are available so the sum of carried traffic will be minimal. Each design will have the same number of overlay switch chassis but the number of connections per switch will be different. The design with logical links (design strategy 2) will have the minimal number of connections because the overlay switches will handle only local traffic flows and no pass through traffic flow, so the switch chassis will be of the smallest size possible. For this 9-node legacy network the overlay backbone design strategy creating a full logical link mesh becomes the lower cost design evaluated. The cost function for this legacy network is dominated by the cost of switches including both chassis and port costs, as well as the unit cost of switches being much greater than the unit costs of links, $K_S >> K_L$. Link distances were minimal and therefore had limited impact upon the total cost. For this type of overlay network design where total switch costs are dominant, total network costs tend to be impacted first by switch costs. Once switch costs are minimized then link costs need to be addressed. A centralized node location with one switch that has high connectivity is the lowest cost overlay design evaluated for this legacy network (Figure 5). As the overlay design strategy includes switches at more legacy nodes, the total switch cost increases and therefore the total design cost increases. The efficient frontier for this legacy topology connects the low cost one-switch scenario, which for this study is the lowest cost design; to the intermediate distribution design and then to the full mesh logical link design.

III.3 North American and NSFNet Case Studies

The North American, NA, and the NSFNet models have network characteristics of longer link distances, average of 437 miles and 1085 respectively (Table 1). For the implementation of these network designs the unit cost of link capacity was set to 1, $K_L = 1$. While the unit cost of switches, K_S, is much greater than the unit cost of links, K_L, long link distances result in the dominant factor being the cost of links. The total link costs for the NA overlay designs evaluated range from $30 Million to less than $80 Million while total switch costs were in the range of $5 Million to less than $30 Million. For the NSFNet case study the link cost range was between $10 and $40 Million and switch cost range was between $1 and $6 Million. The key to minimizing costs for these case studies was to focus on minimizing link costs.

For both these network case studies, the fully distributed switch approach with a full logical link mesh overlay and a switch at every node following backbone design strategy 2 is the most cost effective overlay design evaluated (Figure 6 and 7). The cost structure used in these experiments, coupled with the long link distances, tilts the low cost advantage towards highly connected overlay networks. A full mesh overlay that uses the shortest path connections between each node pair has the smallest possible total link capacity requirement of all possible network designs [Cahn, 1998]. With this cost structure, traffic flow paths need to be minimal for link costs to be minimized. In these experiments, all logical links on a full mesh overlay network are routed over the shortest legacy network path.

For the NA designs evaluated in this study, the designs with overlay switches at most or all nodes tend to be the lower cost. The key to this trend may be both the long link lengths and the low connectivity of the nodes of the legacy NA and NSFNet designs. The low nodal connectivity does not allow many path choices thus total path lengths between overlay designs may be similar. The dominant link costs in this model due to the longer link lengths must be minimized for total cost to be minimized. While the lowest total cost designs evaluated for this study for both the NA case study and the NSFNet were the lowest link cost they were not the lowest switch cost indicating

Figure 6. North American Legacy Overlay Design studies. All of the five growth stages are presented but the efficient frontier is drawn only for the 10X growth period. Lowest cost design strategies evaluated for this study are the distributed overlays with switches at all or most nodes

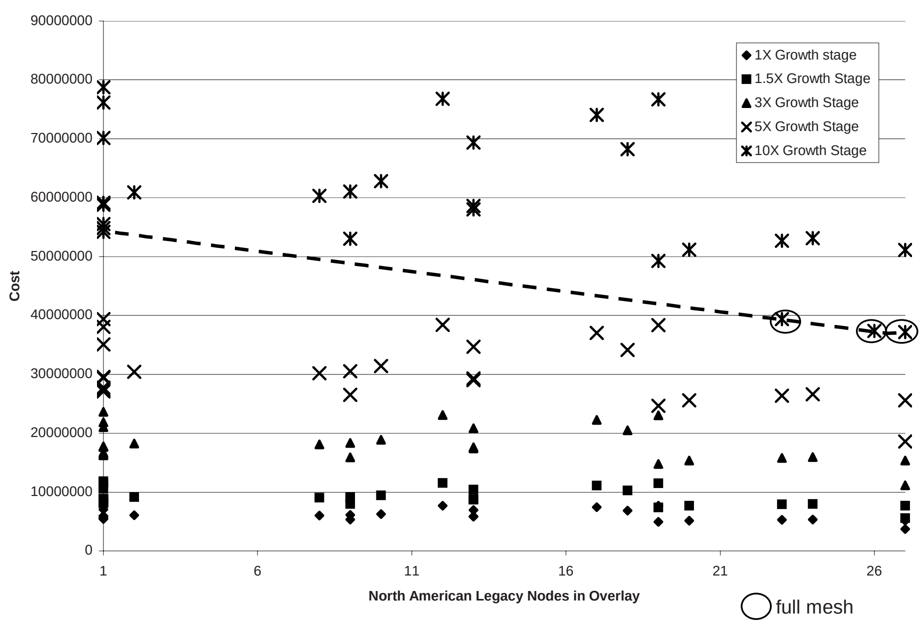
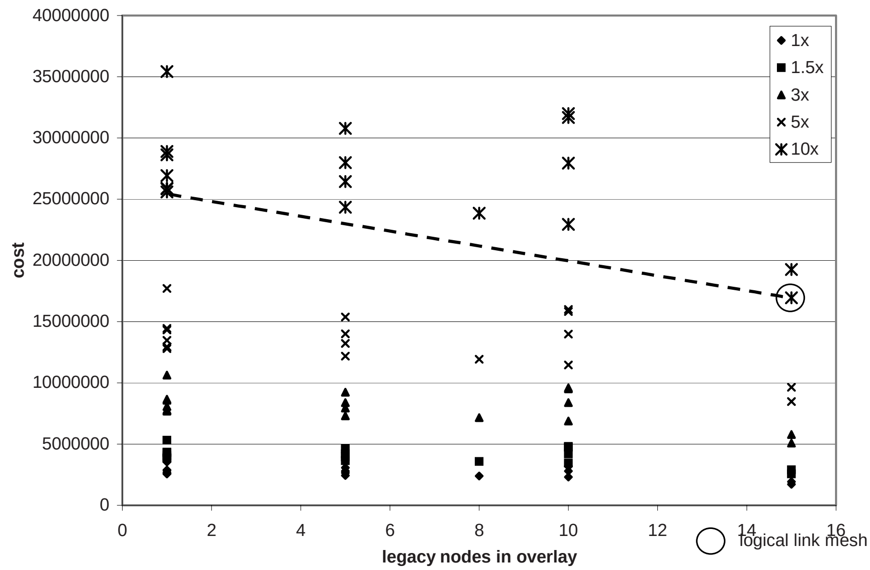


Figure 7. NSFNet Legacy Overlay Design studies. All of the five growth stages are presented but the efficient frontier is drawn only for the 10X growth period. Lowest cost design strategies evaluated are the distributed overlays with switches at all or most nodes



that link costs drove total cost with this legacy topology and the factors used in this study.

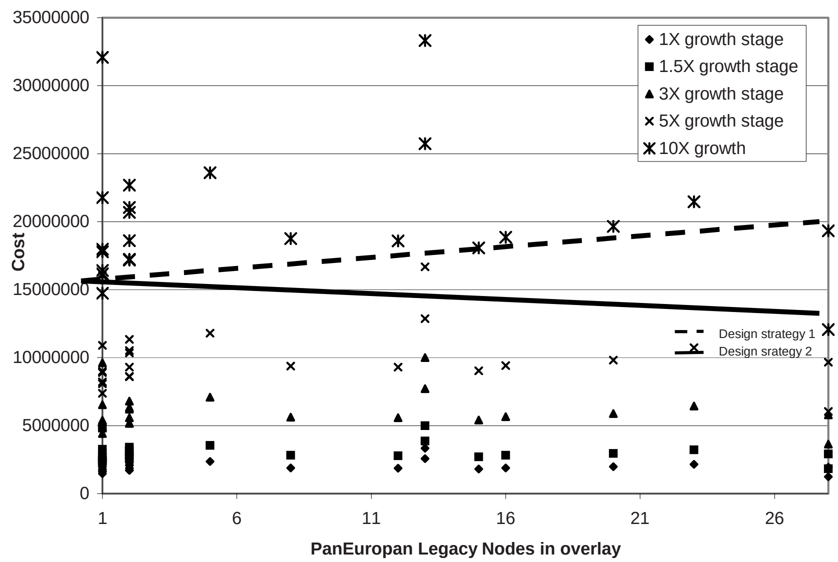
The efficient frontiers, suites of best designs evaluated in this study, for the designs are approximately linear decreasing from one switch to switches fully distributed at all nodes in the legacy network for the NA and NSFNet case studies (Figure 6 and 7). As previously mentioned, in the NA and NSFNet case studies link costs tend to dominate the cost function. For the NA and NSFNet topologies increases in switch costs are offset by decreases in link costs. Increased link costs in more centralized one-switch overlay designs are due to increased amounts of backhaul of traffic flows. These increased link costs offset the decrease in switch costs with fewer overlay switches.

III.4 Pan-European Model Case Study

Key topology characteristics of the Pan-European, PE, network are shorter link distances than NA or NSFNet and more nodes with high connectivity (Table 1). For the PE model network designs, the unit cost of link capacity was set to 1, $K_L = 1$. The unit cost for switches, K_S, was the same as defined for the NA and the NSFNet models, so $K_S >> K_L$. For the designs evaluated in this case study, the range of switch costs was from $3 to 15 Million while the range for link costs was $10-30 Million. The differences between switch and link costs of the designs evaluated in this study are much less than the other two larger case studies and in some instances actually are very similar. Among the designs evaluated for this research, the full logical mesh overlay was again the lowest cost design (Figure 8).

The one-switch overlay approach was the next most cost effective design evaluated in this study but the location of the overlay switch was critical (Figure 8). A central location of high connectivity was important in lowering the total cost of the overlay. There are several possible reasons for this. First, the location of the switch at a centralized node in the legacy network is important because link distances tend to be minimized which in turn minimizes backhaul costs. Periphery locations for a switch require longer flow paths thus increasing

Figure 8. Pan-European Legacy Network overlay design studies. Five traffic demand growth periods for each design scenario were modeled. The efficient frontier is drawn only on the 10X data. The solid line indicates the efficient frontier for the designs with full logical link mesh overlays, design strategy 2. For design strategy 1 that utilized the direct links of the legacy topology the efficient frontier is a dashed line



the distance traffic flow must be backhauled. The degree of connectivity for the periphery location also has an effect upon the total cost, the higher the connectivity of the switch location, the lower the total cost of the overlay.

For the lower cost one-switch overlay designs evaluated in this study, traffic volume from the node location was not an important factor in influencing total cost. The lowest cost one-switch overlay design had the switch located at node 14 (Frankfurt, see Figure 8). This location ranked 21 out of 28 in population size but was a centralized node location with a degree of connectivity of 5. Periphery locations of similar connectivity to that of node 14 but with higher traffic flow demands showed a somewhat higher total cost. This implies that the volume of traffic flow from any given node, while important to the total overall cost, at least for designs evaluated, was less important than minimizing backhaul distances.

The efficient frontier for this case study when drawn to include the full logical link mesh shows the lower cost designs to have more switches distributed over the full PE legacy network (Figure 8). When using design strategy 1 with all switches on the backbone handling all traffic moving through the node to draw the efficient frontier, the opposite relationship is seen. Centralized switch designs tend to be less costly when using design strategy 1 for the PE legacy network evaluated. Were the design strategy 2 applied to overlay backbones consisting of a smaller number of overlay switches, designs with costs lower than that of the single node designs, but in most cases higher than that of the full mesh, are to be expected. The network designer using the design heuristics, knowledge of the traffic flow matrix, and other undefined model constraints could develop a set of designs that define a more continuous efficient frontier for the problem.

IV. SUMMARY

Defining the lowest cost overlay for a legacy network will depend upon the relative value of link cost to switch cost, the traffic flow demands, and the connectivity of the network. When switch costs dominate, minimizing switch costs is the first key to keeping the total cost down. For other topologies where link costs dominate the number of switches becomes less of a factor in the total cost function than minimizing link costs. Determining the dominant influence in the cost function will be the key to minimizing costs and predicting the shape of the efficient frontier. Efficient frontiers for each case study, based on the designs evaluated, were drawn indicating that for the NA and NSFNet case studies the distributed approach with switches at more nodes tend to be more cost efficient, regardless of whether design strategy 1 or 2 is used. For the PE topology case study, based on the designs evaluated for this study, the two design strategies produce different results. Using design strategy 2 with a full logical link mesh design switches at every legacy node was the lowest cost design. If that design philosophy is not used, then a single switch that is located at a central highly connected node is the more cost effective. Defining Efficient Frontiers for suites of network designs is one methodology to understand the impact of cost changes. Future research will focus on other methods of grouping data to define the "best" solution.

V. REFERENCES

Bertsekas, D. (1998). *Network Optimization Continuous and Discrete Models.* Athena Scientific, Belmont, MA.

Cahn, R. S. (1998). *Wide-Area Network Design: concepts and tools for optimization*. Morgan Kaufman Publishers, San Francisco, CA.

Chang, S. G., & Gavish, B. (1993). Telecommunications network topological design and capacity expansion: formulations and algorithms. *Telecommunications Systems, 1*, 99-131.

Copeland, T. E., Weston, J. F., & Shastri, K. (2005). *Financial Theory and Corporate Policy.* Pearson Addison Wesley, Boston, MA.

Drezner, Z., & Hamacher, H. W. (2002). *Facility Location: Application and Theory.* Springer-Verlag, Heidelberg, Germany.

European city mayors population database. Retrieved May 15, 2005. (http://www.citymayors.com/features/euro_cities.html).

Fare, R., Grosskopf, S., & Knox Lovell, C. A. (1994). *Production Frontiers.* Cambridge University Press, Cambridge, Great Britain.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, 120*, 253-281.

Grover, W. D., & Doucette, J. (2001). Topological Design of Survivable Mesh-Based Transport Networks. *Annals of Operations Research, 106*, 79-125.

Kennington, J. L., Olinick, E. V., & Spiride, G. (2007). Basic mathematical programming models for capacity allocation in mesh-based survivable networks. *Omega*, *35*, 629-644.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities, in ed. Koopmans. *Analysis of Production and Allocation* (33-97). New York, NY: Wiley.

Maesschalck, S., Colle, D., Lievens, I., Picavet, M., & Demeester, P. (2003). Pan-European Optical transport networks: An availability-based comparison. *Photonic Network Communications, 5*, 203-225.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investment*. New Haven, CN: Yale University Press.

Sanso, B., & Soriano, P. (1999). *Telecommunications Network Planning*. Kluwer Academic Publishing.

Sawaragi, Y., Nakayama, H., & Tanino, T. (1985) *Theory of Multipleobjective Optimization*. New York, NY: Academic Press

Yun, Y. B., Nakayama, H., Tanion, T., & Arakawa, M. (2001). Generation of efficient frontiers in multi-objective optimization problems by generalized data envelopment analysis. *European Journal of Operational Research*, *129*, 586-595.

VI. KEY TERMS

Carrier Network: A communications service provider network.

Efficient Frontier: The frontier represents the lower boundary of a suite of optimal designs. The efficiency of any design is the relationship of the final cost of a design to the frontier. Evaluating the distance any other design outcome is from the efficient frontier gives a measure of how efficient the design is as related to the parameters used in the optimization. The process was first defined for investment portfolio management and later expanded to production management operations.

Network Design: An iterative process in which planning, development, topology and realization of a new communications service that meets needs from the perspective of service users and or network operators. The design of actual networks is the product of multidimensional constructs of performance and reliability optimized around a certain factor or factors such as cost or reliability

New Service Overlay: An overlay of new service upon an existing communications network.

Chapter LVIII
Combining Small and Large Scale Roaming Parameters to Optimize the Design of PCS

Mohamed Zaki
Azhar University, Egypt

Salah Ramadan
Azhar University, Egypt

ABSTRACT

The cellular principle is an effective way to guarantee efficient utilization of the offered radio band. Although PCS networks use the cellular principle, the next generation of PCS networks needs more improvements in location management to face the increased number of users. Both an Enhanced Profile-Based Strategy (EPBS) for small-scale roaming and a Caching Two-Level Forwarding Pointer (C2LFP) strategy for large-scale roaming have been proposed. This chapter introduces a model that unites the above two strategies. The idea behind this model is based on applying those two location management strategies and specifying the physical parameters of PCS networks from mobility management's point of view so that the underlying solutions can be more cost effective for location management. An evolutionary method using a constraint Genetic Algorithm (GA) has been used to achieve network parameters optimization. For convenience, we selected the planning problem with an appropriate set of parameters to be treated both genetically and analytically. Thus one can easily verify accuracy and efficiency of the evolutionary solution that would be obtained from the genetic algorithm. For more realistic environments, GA could be used reliably to solve sophisticated problems that combine the small-scale and large-scale roaming parameters of PCS networks. A case study is presented to provide a deep explanation of the proposed integration approach.

1 INTRODUCTION

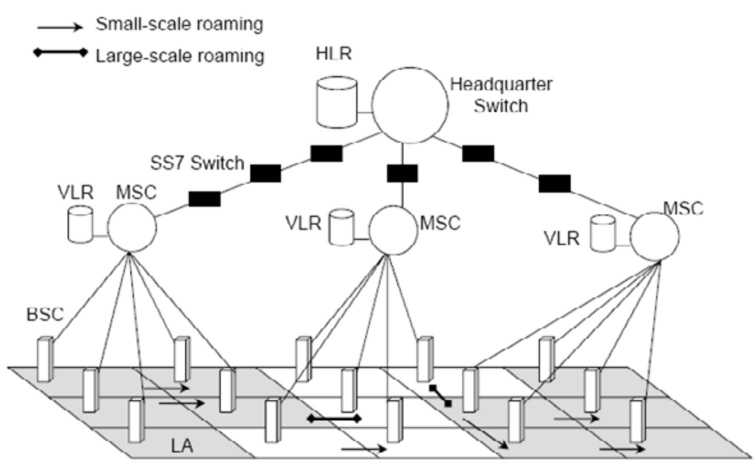
The limited bandwidth for Personal Communication Systems (PCS's) can restrain the growth of the population of mobile customers. In addition, the next-generation of PCS networks should meet the increased requirements in both bandwidth and performance to offer the new demanded services. The bottleneck here is the number of location registrations which leads to a higher demand for switching capacity and signaling links. Definitely, the mobile users must be able to move from one location to another while maintaining access capacity to services regardless of their locations. Therefore, the role of location management in the network is to keep track of Mobile Terminals' (MT's) location information so that calls can be correctly delivered **(Lin, 2003; Lin,2000)**.

Location management is the process of allowing the network to know the attachment point of the mobile user for a successful call delivery. Current location management techniques involve certain database architecture and signaling messages transmission over the PCS network **(Tabbane, 1997; Wong, 2000).** The location management process in a PCS network may use a two-tier system of Home Location Register (HLR) and Visitor Location Register (VLR) databases **(Akyildiz, 1999).** According to these databases, the existing location management scheme allow the mobile user to perform location update at HLR level whenever that user crosses the boundary of an VLR area and deregisters at the previous VLR. Thus, the registrations will generate high signaling traffic as a result of many location updates to the HLR. Since many users are far away from their HLRs, heavy signaling traffic over the network can occur. This problem has been solved in the caching two-level forwarding pointer scheme.

In the case of roaming inside the VLR area, every time the user crosses the boundary of a location area, the existing location management scheme allows the mobile user to perform location update at the VLR level. Thus, the VLR location updates result in high traffic when many users update their locations to the VLR. This problem has been solved in the enhanced profile-based scheme; Figure 1 shows PCS network architecture illustrating two types of roaming. We developed those two roaming solutions where one is for small-scale roaming **(Ramadan S. M., 2004)** and the other is for large-scale roaming **(Ramadan S. M., 2003).** However, the PCS planner can face a problem in developing the locator program using such both solutions in one framework.

Moreover, an evolutionary approach is used to construct this framework, by making use of a GA chromosome, to integrate the small-scale and

Figure 1. PCS Network Architecture



large-scale roaming parameters of PCS networks. Since the combined model involves a constrained optimization problem, the parametric constraints (such as power, setup delay, and list size) are discussed in details. For the reason that the next generation will provide high bandwidth access, there are no bandwidth constraints. The model results are then presented and mapped to physical model network equipments and dimensions are determined. This chapter is based heavily on **(Ramadan S. M., 2003, Ramadan S. M., 2004, Ramadan S. M., 2007-A, and Ramadan S. M., 2007-B)**.

It is organized as follows:

- Background for both of EPBS and C2LFP schemes that will be combined.
- Discussion of the need of integrating those two proposals and description of the significant parameters for the proposed combining model.
- The combination process details.
- The optimization problem.
- A case study for our network design with discussion for the results and mapping the physical network parameters.

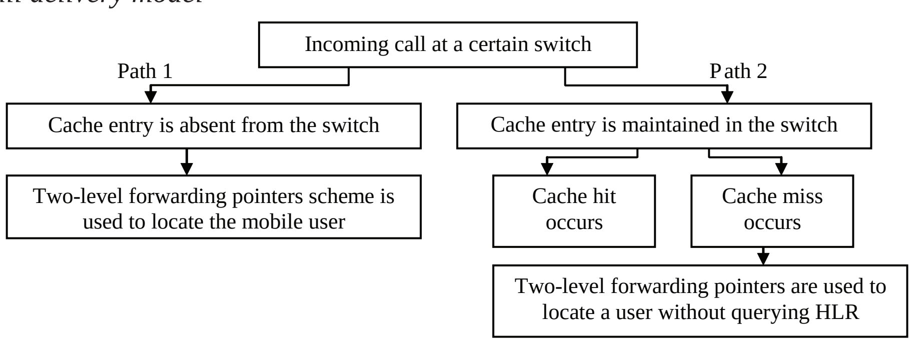
2 BACKGROUND

Location management is one of the most important issues in PCS networks. As the number of mobile users increases, location management scheme under the IS-41 **(EIA/TIA, 1991)** or GSM MAP **(ETSI, 1998)** faces many problems such as increasing traffic in network, and bottleneck occurrence to the HLR. To overcome the entire network problems, two solutions have been proposed: one for small-scale roaming and the other for large-scale roaming.

2.1 Caching Two-Level Forwarding Pointer Scheme:

This scheme is based on exploiting the cache entry stored in a local switch. This switch maintains the recent location of a mobile user when he moves from a Mobile Switching Center (MSC) area, which is served by MSC/VLR, to another MSC/VLR. Whenever the mobile user is accessed through certain switch, an entry contains a mapping from the ID of the mobile user to that of its serving VLR is added to the cache. When a call is initiated for a mobile-user at nearest switch, this calling switch first checks the existence of a cache entry for the mobile user. On one hand, if no entry for the mobile user exists, then the call delivery operation for two-level forwarding pointer scheme is invoked to locate the mobile user; on the other hand: if the mobile user is still residing under the same VLR, a hit occurs and the mobile user is found, and if the mobile has already moved to another location which is not associated with the same VLR, a miss occurs and the call delivery operation, such as two-level forwarding pointer scheme, is invoked to locate

Figure 2. Call delivery model



the mobile user without querying the HLR. Figure 2 describes the call delivery model for C2LFP scheme**.** The two-level forwarding pointer C2LFP strategy is accurately described in **(Ramadan S. M., 2003).**

2.2 Enhanced Profile-Based Scheme

This strategy develops a profile-based scheme (PBS) that is considered a location area (LA)-based strategy; LA is a piece of VLR region. The LAs, that the Mobile Terminal (MT) is most likely to be present in, are registered in a data structure called User Information Table (UIT). The UIT is manipulated to create the main list that is registered to the network during a location update (one copy in the MT and another copy in the network database). If the behavior of a mobile user is changed, then the profile should be modified. Procedures manipulating the UIT will modify the main list after reaching predefined interval.

The modified UIT is sent through certain registration message from the MT to the system. However, the system overwrites the profile using the modified version. Each record in the main list contains three fields, the first two fields are same as in the PBS scheme: the first is for the LA ID, and the second deals with the probability of a user being in this location area a_i, and the last field defines the User State ***US*** which specifies the likelihood of this LA to be present in, the state field has one of two values {*1* or *2*}; by this additional field, a virtual list is created contains the LAs that are the MT is more likely to be present and marked with state 2. This virtual list, called favorite sub-list, stores the LAs that the MT is more expected to be present in his/her daily routine as the work and home LAs.

The ***US*** value is only sent to the network to notify if the MT follows the LAs that belong generally to the main list or especially to the favorite sub-list. The proposed strategy does not need to send a copy of the modified profile in each registration message, but it can send after agreed interval. The advantage of EPBS strategy is to allow the number of LAs stored in the main list to be increased, thus reduces the cost of location update. Furthermore, the existing of multiple states of the MT reduces the cost of paging. The total cost of location update and paging procedures is obviously decreased in the MTs that have relatively constant mobility pattern. The cost computation of the EPBS strategy will be used later. In EPBS strategy, we assume that all of location areas in the list are under the same MSC/VLR; and if the MT moves to a new LA within another MSC, then the system follows the location management method for large-scale roaming (C2LFP scheme). The EPBS strategy algorithm is divided into location update and paging procedures. The algorithms of location update and paging are illustrated in **(Ramadan S. M., 2004).**

3 THE NEED OF COMBINATION

The question here is how those two solutions can run simultaneously, and what would be their combined results. By integrating these solutions in one framework, the network planner may get one coherent model that can answer the aforementioned questions and can find out the physical dimensions of the PCS network. Also, the proposed model possesses optimization features to help in determining a cost-effective location management scheme. In what follows, our model parameters and constraints are illustrated.

3.1 Network Dimensions Parameters

This group includes only static parameters depending on the fluid model to demonstrate the network dimensions parameters **(Thomas R., 1988).** The fluid model characterizes comprehensive movement behavior like the flow of a fluid. Mobile users are assumed to move at an average velocity of ***v,*** and their direction of movement is uniformly distributed over $[0,\pi]$. Assuming that the mobile users are uniformly populated with a density ρ, the location area boundary is of length ***L,*** and the number of crossings through LA per unit time ***U*** is used in our modeling and is given by:

$$U = \frac{\rho v L}{\pi} \qquad (1)$$

Assume the number of zone groups is ***n***, the number of crossings $\mathbf{U_i}$ for zone group ***i*** (where i = 0, 1, 2, …., n) depends on the velocity $\boldsymbol{v_i}$ and density ρ_i of mobile users in this zone, and the perimeter $\boldsymbol{L_i}$ of an LA in this zone where $\boldsymbol{L_i} = 2\pi \boldsymbol{R_i}$. The perimeters $\boldsymbol{L_i}$ or the radii $\boldsymbol{R_i}$ are the network dimensions parameters. Then $\boldsymbol{U_i}$ is obtained from

$$U_i = \frac{\rho_i v_i L_i}{\pi} = 2\rho_i v_i R_i \qquad (2)$$

The small-scale roaming depends on the dimensions parameters because those parameters dynamically influence the rate of location updates for mobile users. It is important to note that the network dimensions parameters R_i should be determined carefully to design the physical model.

3.2 Small-Scale Roaming Parameters

The MSC area is actually emerged into four zone groups where each zone group has certain mobile user velocity and density parameters. The weight λ_i of each zone group ***i*** is obtained from

$$\lambda_i = \frac{U_i}{\sum_{i=1}^{4} U_i} \qquad (3)$$

Where $\boldsymbol{U_i}$ represents the number of crossings of each zone group ***i***. In fact, the rate of location update per user $\boldsymbol{R_L}$, for each zone group ***i***, depends on the network dimensions parameters and can be obtained from

$$R_{Li} = \frac{U_i}{Num_i} \qquad (4)$$

Where $\boldsymbol{Num_i}$ represents the number of mobile users in zone group ***i***, thus from (2)

$$R_{Li} = \frac{4v_i}{L_i} = \frac{2v_i}{\pi R_i} \qquad (5)$$

The total cost of location management for mobile users follow the small-scale roaming (non-traveling users) will be discussed later in this chapter.

3.3 Large-Scale Roaming Parameters

The high mobility users (traveling users) category can be classified by mobility traces which record actual movement behavior for certain segments of the population. In **(Tang D., 2000)**, a seven-week trace has been analyzed the *Metricom metropolitan-area packet radio wireless network*. The results indicate that the locations, which mobile users frequently visit, are close to each other. In addition, the users' movement pattern is Gaussian distributed around the radius of the network. The *Stanford University mobile activity traces* **(Cox D., 1998)** is a publicly available trace generator which includes simulation data of the *Bay Area* location information traces for voice calls.

The large-scale roaming study has two types of parameters: static and dynamic. Some of static parameters depend on the traveling user mobility, and then the value of those parameters is determined by the results of tracing. These parameters are illustrated as follows:

- p_i is the call to mobility ratio for each user group i
- σ_i is the ratio between the cost of FIND operation in case of cache-hit and the cost of basic FIND operation for each user group *i*
- ρ_i is the probability that the cache entry is absent for each user group *i*
- q_i is the probability that the cache-hit occurs for each user group *i*
- K_i is the ratio between the cost of level-1 and level-2 forwarding pointers for each user group *i*
- N_i is the number of incoming calls per time unit for each user group *i which* are determined using small-scale roaming study.

Other static parameters are independent such as

- *S2* The cost of setting up a forwarding pointer (level_2 pointer) between VLRs during caching TwoLevelFwdMOVE.
- *T2* The cost of traversing a forwarding pointer (level_2 pointer) between VLRs during caching TwoLevelFwdFIND
- *m* The cost of basic MOVE operation
- *F* The cost of basic FIND operation

3.4 Physical Constraints

The physical constraints play an important role in determining the suitable values of the system parameters. These constraints are considered the barriers that prevent the system parameters from taking invalid values. Hence, three constraints, excluding the bandwidth, are suggested. In fact, the bandwidth constraint has not been taken into consideration since the spectral efficiency of a wireless network is measured by bits/second per Hertz per cell (BHC), and such efficiency can be improved through frequency reuse and higher-order modulation **(Boleskei H., 2001)** which are beyond the scope of this chapter.

3.4.1 Power Constraint

According to radio wave propagation theorem the strength of the waves decreases as the distance between the transmitter and receiver increases **(Rappaport, 2002),** propagation model has usually focused on predicting the average received signal strength. This model, that predicts the mean signal strength for an arbitrary transmitter-receiver separation distance, is helpful to estimate the ratio coverage area of a transmitter; moreover, it characterizes signal strength over large separation distances (several hundreds or thousands of meters). Also, the free space propagation model predicts received power decays as a function of the separation distance (called the radius of LA in our model). Assume a receiver antenna is separated from a base station by a distance ***R***. Then, the free space power received is given by (6).

$$P_r(R) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 R^2 S} \tag{6}$$

where P_t is the transmitted power, $P_r(R)$ is the received power, G_t is the base station antenna gain, G_r is the receiver antenna gain, S is the system loss factor not related to propagation ($S \geq 1$), and λ is the wavelength in meters. The received power $P_r(R)$ at any distance $R > R_0$ may be predicted from (6), or may be measured in the radio environment by taking the average received power at many points located at a close-in radial distance R_0 from the transmitter. The reference distance must be chosen so that it lies in the far-field region, and R_0 is chosen to be smaller than any practical distance used in the mobile communication system. From (6), the received power in free space at a distance greater than R_0 is given by

$$P_r(R) = P_r(R_0)\left(\frac{R_0}{R}\right)^2 \tag{7}$$

Then

$$R = R_0\sqrt{\frac{P_r(R_0)}{P_r(R)}} \tag{8}$$

If the received power decreases the radius of LA increases. The received power should not decrease under the level it can not be detected. If the threshold of the received power is P_{TH}, then the radius of LA i is constrained by

$$R_i \le R_0\sqrt{\frac{P_0}{P_{TH}}} \tag{9}$$

Each zone in MSC area for small-scale roaming has a different close-in distance R_0, and the radius of LA in each zone R_i is constrained according to (9).

3.4.2 List Size Constraint

This constraint controls the number of LA registered in the list and the number of LAs in a virtual list (favorite sub-list). In **(Ramadan S. M., 2004),** the relation between the cost of small-scale location management and the number of LAs was discussed. The study showed that the list size can be increased up to certain threshold point since after this point the cost of location management is not acceptable. Consequently, the list size must be considered in optimizing the cost of location management in the PCS network.

3.4.3 Setup Delay Constraint

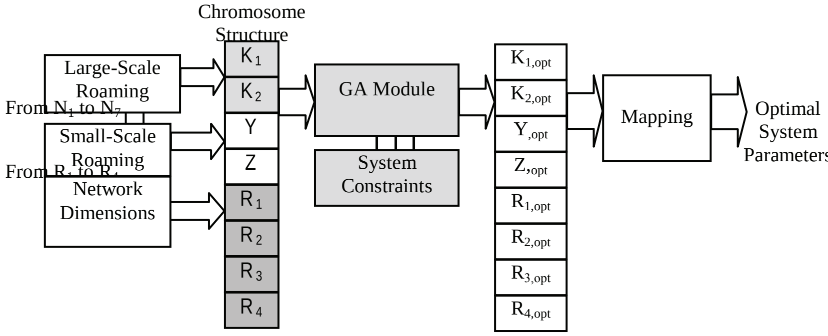
This constraint controls the length of forwarding pointer chain from Mobile Agent/VLR to another agent and the length of forwarding pointer chain from VLR to VLR **(Ramadan S. M., 2003)**. For high mobility users, the longer the chain length is the more reduction in the total cost of large-scale roaming, but the more setup delay. Thus the total chain length must be chosen large enough to accept the setup delay.

4 THE COMBINATION

An evolutionary model is proposed to integrate, the network dimensions, small-scale roaming and large-distance mobility management. Because of numerous interactions between the system parameters and constraints, a genetic-based algorithm is used with the advantage of clarity and robustness. Figure 3 shows the proposed model and illustrates its parameters.

An initial chromosome structure is passed to the genetic algorithm module, and then after

Figure 3. Combined model diagram



running the module under certain constraints, the optimal chromosome containing the optimal system parameters is generated. The results of combining scheme rely upon a global combinatorial optimization process, where all the system parameters are combined in one frame represented by the chromosome.

4.1 Process Nature

The basic idea is to use genetic operators that take constraints into account. An operator can take constraints into account in two ways **(Eiben A. E., 1995; Kielmann T., May 1999)**.

i. It ensures that some of the prescribed constraints always hold
ii. It directs the search by applying heuristics in such a way that the rest of the constraints will be satisfied.

The constraint satisfaction problem (CSP) with ***N*** variables and ***n*** constraints can comprise n <=N domains. Actually, a GA must be modified to be able to cope with CSP. The essential modification is applied on the representation and the child generation mechanism. One has to replace the classical random mechanism of a genetic operator with a mechanism, which takes some of the constraints into account. Since, the Constrained Optimization Problem (COP) is a mixture of free optimization and CSP, then for one particular CSP several equivalent, COP's can be defined by choosing the subset of underlying constraints and/or defining the objective function that measures the violation of the remaining constraints. A genetic algorithm is applicable for COP if **(Eiben A. E., 1995)**:

i. The individuals are elements of the free search space
ii. The fitness function is a monotonic transformation of the objective function
iii. The genetic operators do not lead out of the search space
iv. If the n^{th} population consists of allowed candidates then so does the $(n+1)^{th}$

What makes it difficult to define an applicable GA for a COP is clearly item (iv). This difficulty is mitigated by correctly maintaining the problem constraints in the underlying representation, operators and fitness.

4.2 THE EPBS CALCULATIONS

The total cost of the EPBS scheme S_{EPBS} can be expressed as the sum of costs S_i for location update and paging in each zone *i*.

$$S_{EPBS} = \sum_{i=1}^{n} \lambda_i(\rho, v, R) \times S_i(v, R, Y, Z) \quad (10)$$

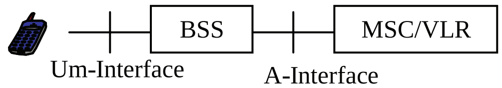
Note that, λ_i represents the weight of each group, and the cost of location update and paging in the i^{th} zone per unit time represents the total cost for small-scale roaming (in-bytes) and is obtained from:

$$S_i = C_L(\frac{2v_i}{\pi R_i})\left[1+\delta_i(\psi_i\left[\frac{1}{\sqrt{Y}}-\frac{1}{\sqrt{Z}}\right]+\frac{1}{\sqrt{Z}}-1)\right]+ 0.64\times C_P(\frac{2v_i}{\pi R_i})\left[1+\delta_i(\psi_i\left[E(Y)-E(Z)\right]+E(Z)-1)\right] \quad (11)$$

The cost S_i depends on two types of parameters: static and dynamic. Some of static parameters depend on the attributes of zone configuration as follows:

- δ_i : the probability that non-traveling user follows the main list.
- ψ_i : the probability that non-traveling user follows the favorite sub-list.
- v_i : the average velocity of mobile users roam through zone i.

Figure 4. Interfaces of PCS networks



Other static parameters are independent such as

- C_L : the cost of one location update process per unit time.
- C_P : the cost of one paging process per unit time.

The cost of C_L and C_P depends on the size of messages (in-bytes) exchanged across the interfaces of the PCS networks. Following **(Tabbane, June 1995)**, the byte unit is chosen to be the cost unit. For non-traveling users, some interfaces are accessed as shown in Figure 4. Those interfaces have been discussed in details in **(Meier-Hellstern, June 1992).**

From **(Tabbane, June 1995)**, *Table 1* shows the number of messages over each interface and the size of message. Then, the cost of C_L is 304 bytes and C_P is 780 bytes. These values reveal why the ratio between two costs is approximately equal to 10. Both of the expected number of paged LAs until the user is found in the main list *E(Z)* and the expected number of paged LAs until the user is found in the favorite sub-list *E(Y)* depend on the fact that users are distributed uniformly or linearly.

The dynamic parameters, that are considered small-scale dynamic parameters and will be processed by genetic algorithms, are shown as follows:

- Z : the size of list that contains all LAs
- Y : the size of favorite sub-list that contains the favorite LAs

Also, the genetic algorithm deals with the dimensional parameters as follows:

Table 1. The Messages Size of Location Management Processes (in-bytes)

Process	**Items**	**Um**	**A**
Location Update (establish and release)	number of messages	6	2
	message size [in-bytes]	44	20
	total cost [in-bytes]	264	40
Paging	number of messages	1	1
	message size [in-bytes]	21.5	11
	total cost [in-bytes]	21.5	11

- R_i : the radius of LA that belongs to zone *i*

4.3 The C2LFP Calculations

The traveling users are classified into:

- High mobility and high usage users as businessmen in train.
- High mobility and low usage users such as the train or high way bus drivers
- Moderate mobility and high usage users such as users in summer trips
- Moderate mobility and moderate usage users
- Low mobility and High usage users
- Others

If the number of types of traveling users is *m* and the weight of each type *i* is w_i, then the total cost of caching two-level forwarding pointer S_{C2LFP} per time unit depends on the sum of costs C_i for MOVE and FIND operations of each user group *i* between two consecutive calls. Then the total cost (in bytes per hour) is obtained from:

$$S_{C2LFP} = \sum_{i=1}^{n} w_i \times N_i(\rho, v, R) \times C_i(k_1, k_2) \qquad (12)$$

Where N_i represents the number of consecutive calls in time unit; actually, it depends on small-scale roaming parameters (see Box 1.).

Box 1.

$$C_i = \frac{S_2}{p_i} + \frac{S_1 - S_2}{(1+p_i)^{k_2} - 1} + \frac{m - S_1}{(1+p_i)^{k_1 k_2} - 1} + \rho_i[F_{hit} + (q_i - 1)F] +$$
$$(1-\rho_i q_i)\left[F + \frac{T_2}{p_i} - \frac{T_2 k_1 k_2}{(1+p_i)^{k_1 k_2} - 1} + \frac{(T_1 - K_2 T_2)\times\left((1+p_i)^{k_1 k_2} - k_1(1+p_i)^{k_2} + k_1 - 1\right)}{\left((1+p_i)^{k_1 k_2} - 1\right)\times\left((1+p_i)^{k_2} - 1\right)}\right] \tag{13}$$

Since $S_1 = K\, S_2$, $T_1 = K\, T_2$, and $F_{hit} = \sigma_i\, F$, then:

$$C_i = \frac{S_2}{p_i} + \frac{S_2(K-1)}{(1+p_i)^{k_2} - 1} + \frac{m - S_2}{(1+p_i)^{k_1 k_2} - 1} + \rho_i[(\sigma_i + q_i - 1)F] +$$
$$(1-\rho_i q_i)\left[F + \frac{T_2}{p_i} - \frac{T_2 k_1 k_2}{(1+p_i)^{k_1 k_2} - 1} + \frac{(K - K_2)T_2\left((1+p_i)^{k_1 k_2} - k_1(1+p_i)^{k_2} + k_1 - 1\right)}{\left((1+p_i)^{k_1 k_2} - 1\right)\times\left((1+p_i)^{k_2} - 1\right)}\right] \tag{14}$$

The cost C_i depends on two types of parameters: static and dynamic. Some of static parameters depend on the traveling user mobility, and then the value of those parameters is determined by the results of tracing. The values of independent parameters *m*, *F*, S_2 and T_2 are determined in **(Meier-Hellstern, June 1992)**, where the message size (in-bytes) sent from one VLR to another is 896 bytes. See Figure 5, for more details.

Table 2 illustrates the volume of traffic messages in large-scale roaming, which there is a message from an MSC associated with new VLR and new VLR itself with load 406 bytes. Afterwards, a message from new VLR to HLR with loads 182 bytes and a message from HLR to old VLR with load 95 bytes are illustrated. Eventually, there is a message from old VLR to new VLR with loads 213 bytes. The total cost for this trip is 896 bytes.

Figure 5. SS7 Location Registration Traffic

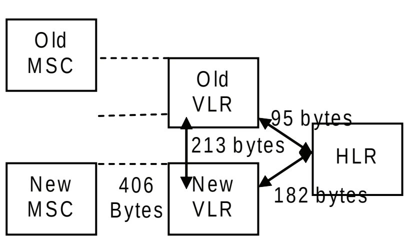


Table 2. Messages size of location registration (in-bytes)

Route Number	Source	Destination	Message Size [bytes]
1	New MSC	New VLR	**406**
2	New VLR	HLR	**182**
3	HLR	Old VLR	**95**
4	Old VLR	New VLR	**213**
			Total = 896

Both of Basic MOVE and Basic FIND operations need to do this trip twice to complete their procedure, where m = F = 1792 bytes. In addition, S_2 = 812 bytes, and T_2 = 406 bytes. The dynamic parameters that are considered dynamic large-scale parameters will be processed by genetic algorithms are shown as follows:

- k_1 : the threshold of level_1 pointer chain in large-scale roaming
- k_2 : the threshold of level_2 pointer chain in large-scale roaming

5 THE OPTIMIZATION

The problem of PCS network planning is generally intractable because it is not easy to move from a starting point towards a better situation. When one factor (parameter) is changed, other factors are consequently changed. Actually, intractable realistic planning and/or design problems can not be subject to analytical treatment. In this case, genetic algorithms (GA's) may be used with the advantage that:

- They tackle the realistic intractable problem of planning PCS networks that may include discontinuous objectives.
- They can emphasize the parameters that are statistically significant and physically meaningful.

5.1 Chromosome Construction

Generating a correct chromosome is initially needed for starting the GA. The chromosome must contain the dynamic parameters in the proposed location management schemes as follows:

- k_1 : the threshold of level_1 pointer chain
- k_2 : the threshold of level_2 pointer chain
- Z : the size of list that contains all visited LAs under MSC area
- Y : the size of favorite sub-list that contains the favorite visited LAs through the main list
- R_i : the radius of LA that belongs to zone *i*

5.2 The Fitness Function

The fitness function is needed to guide GA search, which the goal of the combined model is to maximize the value of the fitness function. Our problem is optimizing the cost of location management in PCS network that uses the proposed location management the schemes. This optimization occurs when minimizing the total cost. The total cost (in bytes per hour) is obtained from:

$$Total\_Cost = Cost_1 + Cost_2 \tag{15}$$

Cost_1 is the cost of EPBS scheme and *Cost_2* is the cost of C2LFP scheme. Since the criteria function must be maximized, and then it is reciprocal of the total cost as follows:

$$Fitness = \frac{1}{Total\_Cost} \tag{16}$$

5 CASE STUDY

To implement the combined model, we have relied upon a well-known geographical area as a case study to verify that the proposed location management schemes are effective compared with the existing scheme. For large-scale roaming, firstly, some measurements will be taken from a public communication network (e.g., Metricom which covers San Francisco Bay Area, Washington D.C., and Seattle). Secondly, the measurements resulted from the mobility traces are exploited in the caching two-level forwarding pointer solution. The fluid model parameters, for the enhanced profile-based strategy, are functions of the users' densities and velocities. Also, the influence of the network dimensions is taken into consideration in small-scale roaming.

5.1 The Use of Fluid Model

The fluid model is used in modified profile-based schemes and characterizes aggregate movement behavior as the flow of a fluid **(Thomas, Sep. 1988).**

The rate of users, moving out of LA, depends on the velocity of users and area density. Thus, one can classify the cost items of the EPBS scheme into 4 categories according to the type of zone. Each zone, which represents the covered area by VLR/MSC, has radius 100 km and is divided into 4 categories, Figure 6:

- City Center with radius 10 km
- Urban area with radius 20 km
- Rural area with radius 30 km
- Desert Roads with radius 40 km

There are three categories of users in each area zone covered by VLR/MSC:

1. Working people
2. Residential people
3. High mobility people

Table 3 shows the percentage of each category in the zone and the percentage of vehicle used in each category. Each category can move from LA to another inside certain zone by using private car, taxi, or public transportation.

Table 4 shows the density of users in each category, and the speed of vehicle in each region that of course is inversely proportional to users' density.

5.2 Mobility Traces

The mobility tracing method is applied to *Metricom* network as a case study, which a seven week trace analysis is used. This trace finds how users take advantage of a mobile environment [11], also finds answers to overall network questions such as when the mobile network is active, how active the network gets, where the network is active, as well as radio mobility questions such as how far, how often, and when users move. The answers to such questions are crucial in planning the mobile network infrastructure, and in understanding how people actually take advantages of a mobile network.

The mobility traces method is based on Metricom data, where Metricom has installed a Ricochet TM packet radio network infrastructure in three major metropolitan areas (San Francisco Bay Area, Washington D.C., and Seattle), as well as in some airports, hotels, and college campuses scattered across the United States. To summarize the results of this study, Table 5 shows the necessary data for caching two-level forwarding pointer module that clusters the results into 11 categories.

Figure 6. Chart of Coverage area for VLR/MSC

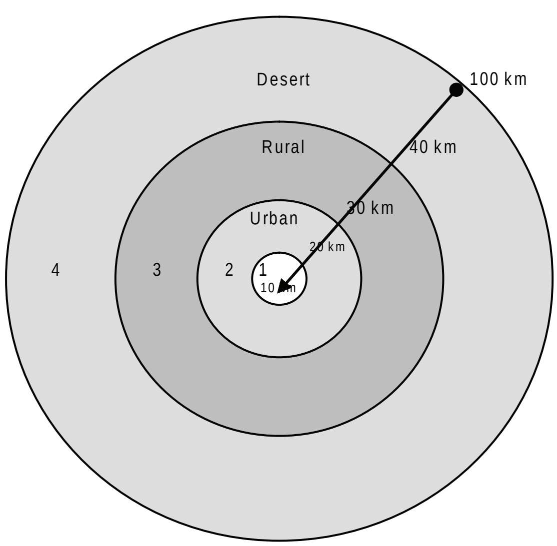


Table 3. Categories of users that belong to zone, and vehicle types that are used by each category

Working people	60%	Private Car Taxi Public Transportation	35% 5% 60%
High mobility users	25%	Private Car Taxi	30% 70%
Residential people	15%	Private Car Taxi Public Transportation	18% 16% 66%

Table 4. Types of regions in area zone, radius of the region, capacity of each region, and the speed of vehicles in each region

Region	**Radius**	**user/ km²**	**Vehicle speed**	
			Public Trans.	**Car/Taxi**
City Center	10 km	5000	5-15 km/hr	10-20 km/hr
Urban	20 km	2000	10-30 km/hr	15-40 km/hr
Rural	30 km	500	30-70 km/hr	40-80 km/hr
Desert roads	50 km	100	60-80 km/hr	60-100 km/hr

Table 5. Cumulative Data of 11 user categories

Num	User Type	P	ρ	q	σ	*K*	N	w
1	Stationary	∞	-	-	0.3	-	2.1	18.6%
2	Moderate mobility Moderate usage	2.55	0.7	0.8	0.3	2	4	6.47%
3	Moderate mobility Low usage	2.1	0.6	0.7	0.3	2	2.67	5.74%
4	Minimal mobility Moderate usage	1.7	0.6	0.7	0.3	2	4	19.23%
5	Moderate mobility High usage	1.43	0.5	0.7	0.4	3	4.33	3.98%
6	High mobility Moderate to low usage	0.615	0.5	0.5	0.4	3	3	5.89%
7	Moderate mobility Minimal usage	0.467	0.4	0.5	0.4	3	2.5	4.5%
8	High mobility High usage	0.454	0.4	0.5	0.5	4	4.33	5.24%
9	High mobility Moderate usage High connection rate	0.211	0.4	0.4	0.5	4	4	3.07%
10	Moderate mobility Train users	0.1	0.3	0.4	0.6	4	5	3.56%
11	Others	0	0.3	0.4	0.6	4	3.1	23.72%

The larger the average distances are the lower probabilities for ρ and q would be. The resulted values for static parameters are illustrated in Table 5 as cumulative data.

5.3 Model Input

The static parameters for both of large-scale roaming and small-scale roaming are determined respectively. The caching two-level forwarding pointer solution contributes to two dynamic parameters k_1 and k_2 in the chromosome which have some possibilities under constraint $k_1 * k_2 <= 12$ that represents setup delay constraint. The possibilities of k_1 are 1, 2, 3, and 4 respectively. The possibilities of k_2 are 1, 2, 3, 4, 5, 6, 7, and 8 respectively. The enhanced profile-based solution introduces two dynamic parameters *Y* and *Z*, where *Z* represents the number of LAs that are stored in the list, and *Y* represents the number of LAs belonging to favorite sub-list. The possibilities of *Z* are 4, 8, 12, 16, 20, 24, 28, and 32 respectively. The possibilities of *Y* are 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8 of the value of *Z* respectively.

Also, the network dimensions have four dynamic parameters R_1, R_2, R_3, and R_4 represent the radii of city center, urban, rural, and desert regions respectively. The range of R_1 is from 100 to 850 meters with step 50 meters; the range of R_2 is from 0.5 to 2 kilometers with step 100 meters; the range of R_3 is from 1.4 to 4.6 kilometers with step 200 meters; and the range of R_4 is from 2.5 to 10 kilometers with step 500 meters.

5.4 The Genetic Optimization Process

The resulted chromosome consists of 8 fields; each field is encoded by certain number of bits as follows:

R_4 : 4-bits ; R_3 : 4-bits ; R_2 : 4-bits ; R_1 : 4-bits ; Y : 3-bits ; Z : 3-bits ; K_2 : 3-bits ; K_1 : 2-bits

Thus, the length of the chromosome is 27 bits. The genetic algorithm is guided by a fitness function to determine the fitness of each chromosome in the population. Therefore, the chromosomes with maximum fitness are chosen to create the next generation. The fitness function is modeled as follows:

$$\textbf{\textit{Maximize:}}\ Fitness = \frac{1}{Net\_Cost} \quad (17)$$

$$Net\_Cost = Cost_1 + Cost_2 \quad (18)$$

Where ***Cost₁*** represents the cost of the EPBS solution, and ***Cost₂*** represents the cost of the C2LFP solution

5.4.1 EPBS Cost Function

$$\textbf{\textit{Minimize:}}\ Cost_1 = \sum_{i=1}^{4} \lambda_i(\rho, v, R) \times S_i(R, Y, Z)$$

Constrained by:
$Y < Z$
$Y < maximum\_list\_size$

The cost per zone group $S_i(R,Y,Z)$ is determined from (11), and ρ_i is given in *Table 4*. To determine the velocity of users in each region, *Table 4* is also used to obtain the speed of vehicles and the types of users in each region. The probability density distribution can be either linear or uniform. For linear distribution:

$$v_i = v_{\max}\left(1 - \frac{\rho_i}{\rho_{\max}}\right) \quad (19)$$

The *E(Y)* and *E(Z)* for uniform distribution as follows:

$$E(Z) = \frac{Z+1}{2} \quad (21)$$

$$E(Y) = \frac{Y+1}{2} \quad (22)$$

The *E(Y)* and *E(Z)* for linear distribution are as follows:

$$E(Z) = \frac{Z+2}{3} \quad (23)$$

$$E(Y) = \frac{Y+2}{3} \quad (24)$$

5.4.2 C2LFP Cost Function

$$\textbf{\textit{Minimize:}}\ Cost_2 = \sum_{i=1}^{11} w_i \times N_i(\rho, v, R) \times C_i(k_1, k_2)$$

Constrained by:
$k_1 \times k_2 \le 12$

As mentioned previously, the weights of eleven mobility types, which are measured by tracing similar network for seven weeks, are obtained. Also p_i, ρ_i, q_i, σ_i, and K_i are calculated. The cost of Basic MOVE and Basic FIND operations is m = F = 1792 bytes. In addition, S_2 = 812 bytes, and T_2 = 406 bytes.

5.4.3 System Constraints

As mentioned previously, the objective function of combined model has eight parameters; those parameters are affected by the following constraints:

- **Power Constraint:** This constraint controls R_1, R_2, R_3, and R_4. These values represent radii of LA in city center, urban, rural, and desert regions respectively. So that:

$$\frac{P_i}{P_{0i}} >= \left(\frac{R_{0_i}}{R_i}\right)^2 \quad (26)$$

Then

$$R_i <= R_{0i}\sqrt{\frac{P_{0_i}}{P_i}} \tag{27}$$

The P_o represents the reference power for each region, and R_0 represents the reference distance between the transmitter and the receiver for each region.

- **Setup Delay:** This constraint controls the values of k_1 and k_2 with constraint $k_1 * k_2 <= 12$ **(Ma W., Jan. 2002).**
- **List Size:** This constraint controls *Z* and *Y*, where for uniform distribution, the list size must be less than 25, and for linear distribution, the list size must be less than 37.

5.5 Model Output

After running the genetic algorithms to choose the best chromosome for network planning, the next chromosome is obtained; regardless the user distribution is uniform or linear.

R_4 : 0101 ; R_3 : 1000 ; R_2 : 0101 ; R_1 : 0111 ; Y : 011 ; Z : 001 ; k_2 : 101 ; k_1 : 01

After decoding the resulted chromosome, we find that: $k_1 = 2$, $k_2 = 6$, Z = 8, Y = 4, $R_1 = 0.7$ km, $R_2 = 1.4$ km, $R_3 = 4.2$ km, and R4 = 7 km.

5.6 Mapping to the Physical Network Parameters

The corresponding physical model calculates the area of LA and the number of LAs in each zone according to the GA's results. Also, the expected memory is determined for both VLR and HLR. Those issues will be discussed as follows:

5.6.1 Area and Number of LAs

From the modeling of EPBS solution, the radii of city center, urban, rural, and desert regions are 10 km, 20 km, 30 km, and 40 km respectively as shown in Figure 6. In addition, the area of each region can be calculated, and then we can obtain 314 km^2, 2512 km^2, 8478 km^2, and 20096 km^2 for city center, urban, rural, and desert regions respectively. Also, the given radii from chromosome help in obtaining 1.54 km^2, 6.154 km^2, 55.39 km^2, and 153.86 km^2 for LA in city center, urban, rural, and desert regions respectively.

Now, the number of LAs in each region can be calculated as follows:

$$N_1(LA) = \frac{Area(CityCener)}{Area(LA\_in\_CityCener)} = \frac{314}{1.54} \cong 204$$

$$N_2(LA) = \frac{Area(Urban)}{Area(LA\_in\_Urban)} = \frac{2512}{6.154} \cong 408$$

$$N_3(LA) = \frac{Area(Rural)}{Area(LA\_in\_Rural)} = \frac{8478}{55.39} \cong 153$$

$$N_4(LA) = \frac{Area(Desert)}{Area(LA\_in\_Desert)} = \frac{20096}{153.86} \cong 131$$

Expected number of users, that will visit coverage area by one VLR, is obtained from:

$$E = \sum_{i=1}^{4} \rho_i A_i \approx 12.8 * 10^6 \quad Users$$

Consequently, we need $12.8*10^6$ lists in each VLR; also the list size is eight and the number of favorite LAs is four from the resulted chromosome, then each list needs eight records, and each record contains (14-bits): 10 bits are needed for LA ID because the total number of LAs under

MSC area is 896, 3-bits for the LA probability, and one bit to tell whether the LA is favorite or not. Accordingly, each list needs 112 bits or 14 bytes. Then, the e*xpected memory size for one VLR = E * List_Size(in Bytes) = 180 MB*. In this case study, there are 13 VLRs; thus, the expected memory size in the HLR is less than 2.4 GB.

6 DISCUSSION AND FUTURE TRENDS

Since genetic algorithms offer evolutionary, rather than closed form solutions; their results might be verified by applying the classical gradient technique to the network-planning problem. In its general form the problem parameters are nonlinear, discontinuous and interrelated. Therefore, it is not adequate to use the gradient optimization method because most likely it will fail to find the required global optimum solution. For convenience, a simplified network model is obtained by omitting the following two dependencies:

1. The dependency of small-scale roaming on the network dimensions
2. The effect of small-scale motion profiles on long-distance mobility management.

Accordingly, the nonlinear planning optimization problem can be expressed as follows:

Minimize: Net_Cost (k_1, k_2, Y, Z, R_1, R_2, R_3, R_4)

Subject to:

$k_1 * k_2 \leq 12$
$Y - Z \leq 0$
Power Ratio ≤ 0.5
$0 \leq k_1 \leq 12$
$0 \leq k_2 \leq 12$
$0 \leq Y \leq 25$
$0 \leq Z \leq 25$
$100 \text{ m} \leq R_1 \leq 850 \text{ m}$
$500 \text{ m} \leq R_2 \leq 2000 \text{ m}$
$1400 \text{ m} \leq R_3 \leq 4400 \text{ m}$
$2500 \text{ m} \leq R_4 \leq 10000 \text{ m}$

Then, we applied both the proposed GA for constrained optimization and the constrained nonlinear gradient optimization method **(Wiley M., 1995)** on the simplified network model, as such, at various input parameters; the results of all runs have verified the applicability of the proposed GA.

Here the emphasis is how to design an adaptive, flexible, and evolutionary tool to optimize the solution of the mobility management problem by exploiting the competition of the underlying GA chromosomes. In the future, it might be promising to make use of a cooperative rather than our competitive revolutionary genetic algorithm. In this case, the solution procedure could be inspired by a swarm based system. The expected tool will be adaptive, decentralized, flexible and robust. Moreover, it will provide self organization in the sense that it provides mechanism whereby higher-level forwarding pointers are constructed depending upon the interactions raised from the lower level forwarding pointers. The rules, specifying such interactions, are executed on purely local information.

7 CONCLUSION

Location management in cellular networks needs a comprehensive model that can be reliably adopted in network design and planning. This model helps in deciding if the optimal design meets the increased requirements of next generation, such as bandwidth and performance. In this chapter, a combined model integrates network dimensions, profile-based small-scale roaming, and long-distance location management scheme.

To allow natural interactions among parameters; a genetic approach is exploited for integration where each network parameter is represented by a gene. The genes can form a chromosome that comprises a valid solution in the solution space. The chromosome consists of eight genes: two genes represent level-1 and level-2 forwarding pointer chains, two genes represent both the sizes of main list and favorite list in the EPBS, and four genes represent the network dimensions. By applying the genetic operations, guided by a fitness cost function, on a population of chromosomes, a constrained optimization process is carried out to obtain the best values of network parameters.

To verify the effectiveness of the proposed genetic approach, a simplified network model was considered, where both the genetic approach and the nonlinear gradient method have been applied and the optimization results are compared. After that, a straightforward mapping procedure was applied to obtain the physical network parameters from the corresponding optimization model output. The PC network planner can obtain the optimum values of the system parameters such as the optimum values of the location area radii, the number of LA's, and the expected memory for VLR and HLR.

REFERENCES

Akyildiz, I. F., Mcnair, J., Joseph, S. M., & Wang, W. (1999). Mobility Management in Next-Generation Wireless Systems. *Proceedings of the IEEE, 87*(8), 1348-1384.

Boleskei, H., Paulraj, A. J., Hari, K. V. S., & Naber, R. U. (2001). *Fixed Broadband Wireless Access: State of the Art, Challenges, and Future Directions*, IEEE Communications Magazine, (pp. 100-108). Jan, 2001.

Cox, D., & Widom, J. (1998). The Sumatra project has been supported by the Centers for Telecommunications and Center for Integrated Systems at Stanford, by Pacific Bell, and currently by the National Science Foundation. Homepage: http://www-db.stanford.edu/sumatra

EIA/TIA, Technical Report IS-41, Revision B (1991). *Cellular Radio Telecommunications Intersystem Operations.*

Eiben, A. E., Raué, P. E., & Ruttkay, Z., (1995). *How to Apply Genetic Algorithms to Constrained Problems.* In L. Chambers (Ed.), *Practical Handbook of Genetic Algorithms, 1.* CRC Press, Inc.

ETSI, Technical Report GSM, Digital Cellular Telecommunications System, Phase 2+ (1998). *Mobile Application Part (MAP) Specification (GSM 09.02 version 7.51 Release).*

Kielmann, T., & Hofman, R., (1999). *The Genetic Algorithms* homepage: ftp.cs.vu.nl/pub/papers/ai/GA/. E-mail:kielmann@cs.vu.nl.

Lin, Y. B., & Chlamtac, I. (2000). *Wireless and Mobile Network Architectures.* John Wiley & Sons, Inc.

Lin, Y. B, Haung, Y. R., Chen, Y. K., & Chlamtac, I. (2003). Mobility Management: from GPRS to UMTS. *Wireless Comm. and Mobile Computing, 1*(4), 339-359.

Ma, W., & Fang, Y. (2002). Two-level forwarding pointer for location management in PCS networks. *IEEE Transactions on Mobile Computing, 1*(1), 32-45.

Meier-Hellstern, K. S., Alonso, E., & O'Neil, D.R. (1992). The use of SS7 and GSM to support high density personal communications. *In Proc. IEEE ICC'92*, Chicago, IL, paper 356.2.

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2003). Two-Level Forwarding Pointers-Based Cache Scheme for Location management in PCS Networks. *The 46Th IEEE Midwest Symposium On Circuits and Systems*, Cairo, 226C, Dec. 2003.

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2004). *Two-Level Profile-Based Scheme for Location Update and Paging in PCS Networks.* In *Scientific Bulletin of Ain-Shams University, 39*(4).

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2007-A). An Integrated Model to Design PCS Networks. *Wireless Personal Communications*, DOI : 10.1007/s11277-006-9126-6, *41*, 19-41.

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2007-B). Location Management in PCS Networks by Caching Two-Level Forwarding Pointers. *Wireless Personal Communications*, DOI : 10.1007/s11277-006-9128-4, *41*,57-75.

Rappaport, T. S. (2002). *Wireless Communications Principles and Practice Systems*. India, Delhi, Pearson Education, Inc. Asia. 2002.

Tabbane, S. (1995). An Alternative Strategy for Location Tracking. *IEEE J. Selected Areas in Comm.*, *13*(5), 880-892.

Tabbane, S. (1997). Location Management Methods for Third-Generation Mobile System. *IEEE Comm. Magazine*.

Tang, D., & Baker, M. (2000). *Analysis of a MetropolitanArea Wireless Network*. Wireless Networks, Department of Computer Science, Stanford University.

Thomas, R., Gilbert, H., & Mazziotto, G. (1988). Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network. *In Third Nordic Seminar on Digital Land Mobile Radio Communication*. Copenhagen.

Wiley, M. (1995). LindoApi optimization tool. Site: http://www.lindo.com/

Wong, V. W. S., & Leung, V. C. M. (2000). Location Management for Next-Generation Personal Communications Networks. IEEE Network, (pp.18-24).

KEY TERMS

Caching 2-Levels Forwarding Pointers (C2LFP): A solution proposed to minimize the cost of location management for large-scale roaming inside PCS network.

Enhanced Profile-Based Scheme (EPBS): A solution proposed to minimize the cost of location management for small-scale roaming inside PCS network.

Home Location Register (HLR): One of two tire-database systems. It's considered the centralized point of PCS network.

Location Area (LA): Piece of RA supervised by one or more base stations.

Location Management (LM): One of the functions of mobility management layer.

Mobility Management (MM): One of the radio layers of PCS reference model.

Personal Communications Systems (PCS): The major area of research. Also, PCS networks can be called Cellular Networks.

User Registration Area (RA): Piece of the whole coverage area supervised by one VLR.

Visitor Location Registers (VLRs): Collection of nodes supervises the whole network and connected with HLR through SS7 network.

Chapter LIX
Secure Resource Optimization in Distributed Service Computing

Kaiqi Xiong
Texas A&M University, USA

Harry Perros
North Carolina State University, USA

ABSTRACT

This chapter considers a set of computer resources used by a service provider to host enterprise applications for customer services subject to a service level agreement (SLA). The SLA defines three QoS metrics, namely, trustworthiness, percentile response time and availability. We first give an overview of current approaches, solutions and challenges in the security-aware resource optimization problem. Then, we present a framework for solving the problem. We further propose an approach for resource optimization in such an environment that minimizes the total cost of computer resources used by a service provider for such an application while satisfying all these three QoS metrics in a security-aware resource optimization problem that typically arises in distributed service computing. We formulate the security-aware resource optimization problem as an optimization problem under the SLA constraints, and solve it using an efficient numerical procedure. Finally, we conclude our discussion and provide the research directions for future study.

INTRODUCTION TO SECURITY-AWARE RESOURCE OPTIMIZATION

With the number of e-Business applications dramatically increasing, service level agreement (SLA) will play an important part in distributed service computing. An SLA is a combination of several qualities of services (QoS), such as security, performance, availability, and reliability, agreed between a customer and a service provider. The service provider may be a telecommunications carrier, an Internet service provider, or any company that provides outsourcing services. The services provided may include dedicated leased lines, shared packet-oriented services, Web hosting services, off-site application management, and off-site network management. With the ubiquity of mobile devices such as smartphones and PDAs, mobile devices will generate a large percentage of Web service requests.

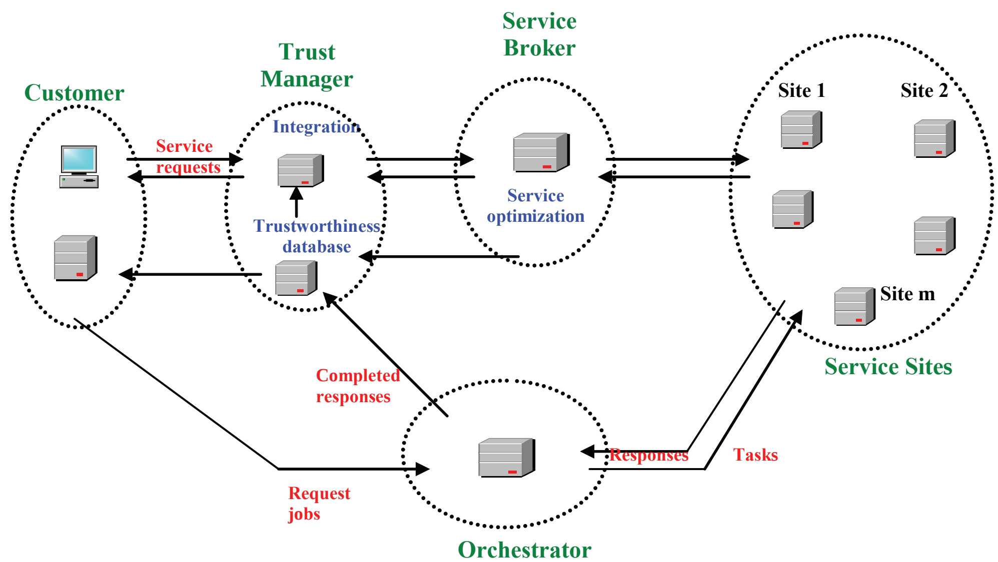
An SLA defines all aspects of the service being provided. It generally consists of security, performance and availability. *Security* can be categorized as *identity security* and *behavior security*. Identity security includes the authentication and authorization between a customer and a service provider, data confidentiality and data integrity. Behavior security describes the trustworthiness among multiple resource sites, and the trustworthiness of these resource sites by customers, including the trustworthiness of computing results provided by these sites. In the chapter, we are only concerned with trustworthiness. *Performance* includes the two following aspects.

- *Response time* is the time for a service request to be satisfied. That is, this is the time it takes for a service request to be executed on the service provider's multiple resource sites. The definition is consistent with the term in queueing theory. That is, it is the total time that a service request spends in a queueing system. We must point out that the definition of response time here is usually known as the *turnaround time* in a computer system. However, we will not be confused by the definition of response time since a discussion of response time in the chapter is restricted in a queueing system. More precisely, we will propose a queueing network method to calculate the response time in the chapter.
- *Throughput* is the service rate that a service provider can offer. It is defined as the number of service requests whose processing is completed in a single unit of time.

Finally, *service availability* is the percentage of time that a service provider can offer services.

In this chapter, we shall discuss security-aware resource optimization in service computing application under SLA guarantees. We first give an overview of existing research taken on this subject, and briefly discuss the challenges of

Figure 1. An SLA-based Web service model



solving the security-aware resource optimization problem. Then, we propose an approach for solving the problem. Specifically, we define and solve a security-aware resource optimization problem that occurs in Web service applications, subject to the constraints of trustworthiness, percentile response time and availability.

Figure 1 depicts a framework for this security-aware resource optimization problem in Web services. In this figure, a customer represents a business that generates a stream of service jobs issued at a specified rate, which require a certain quality of service. A trust manager, who represents the customer, negotiates a SLA with a service broker who represents resource sites (alternatively called service sites) $S_1, S_2, \ldots, S_M$ where M>0. The trust manager checks the trustworthy information of the resource sites and selects m ($0<m \leq M$) say sites 1, 2, …, m of those sites which meet pre-defined trustworthy requirements for serving the service jobs. Then, the service broker optimizes these m service sites in terms of service costs so that the service jobs can receive the requested quality of service. The trust manager monitors the trustworthiness of these service sites and if necessary, it may replace one or more if their trustworthiness index drops below a pre-defined level. Specifically, service jobs are sent directly to the orchestrator who is responsible for initiating a sequence of tasks necessary for the execution of a service job. The result is sent back to the trust manager who forwards it to the customer, after it checks its trustworthiness and updates its database. The customer also sends feedback to the trust manager who uses it to update its trustworthiness information. This framework will be further discussed in detail later. We will formulate the security-aware resource allocation problem as an optimization problem under SLA constraints in which we calculate the number of servers in each service site that minimize a cost function that reflects operational costs for customer services. We will provide a solution to this problem using an efficient algorithm and an illustrative example to briefly demonstrate how to use the algorithm.

The rest of this chapter will mainly focus on the following three aspects: (1) Current approaches, solutions and challenges; (2) Our proposed solution and practical implication; and (3) Conclusions and future directions.

CURRENT APPROACHES, SOLUTIONS AND CHALLENGES

The issue of service quality has been extensively investigated. Characterizing QoS metrics is essential but challenging in the security-aware optimization problem.

The first QoS metric is trustworthiness. "Trust" is used to deal with the notion of the trustworthiness in this research. In this chapter, trust is defined as a firm belief in the competence of a resource site that acts as expected. Many researchers have studied the trustworthiness of service entities, and suggested several different trust metrics. Zhang et al. (2004) and Ziegler & Lausen (2004) classified various trust metrics. Vu et al. (2005) proposed a new QoS-based semantic Web service selection and ranking solution using a trust and reputation management and assuming known QoS qualities. Azzedin & Maheswaran (2002) studied trust management in Grid computing.

Furthermore, availability is a critical metric in today's computer design. A computer system is unavailable due to a variety of causes, such as network, hardware, software, and operational failures, security attacks, and so on. Detecting and preventing these failures and attacks is beyond the scope of our discussion in this chapter. Cisco has asserted that the operational failure causes 80% of non-availability (see CISCO (2004)). Hence, increasing network availability is becoming a key priority for enterprise and service provider organizations. Brown & Patterson (2000) defined a new availability metric to capture the variations of the system quality of service over time. It is defined by the number of requests satisfied per second (or the latency of a request service) and the number of server failures that can be tolerated by a system.

Among the QoS metrics, the response time is the most important one. The computation of the response time has been extensively studied for a variety of computing systems. However, only the average response time is calculated rather than a percentile of the response time in most literature. Levy et al. (2003) presented a performance management system for cluster-based Web service where the average response time is used as a metric. Chandra (2003) employed an online measurement method, and considered a resource optimization problem based on measured response times. In order to compute a percentile of the response time one has to first find the probability distribution of the response time. This is not an easy task in a complex computing environment involving many computing nodes. Xiong & Perros (2006b) developed the efficient and accurate solutions of the percentile of the response time for both a tandem service model and a feedback service model.

Web service is often contracted through SLAs that typically specify a certain QoS in return for a certain price. Although QoS was not defined in the initial UDDI standard for Web services, many studies have been carried out to extend the initial UDDI, such as WSLA (Keller & Ludwig (2003)), WSOL (Tosic et al. (2002)), and QML (Dobson (2002)). The issue of a reputation-based SLA has been studied by Jurca & Faltings (2005) where the cost is determined by the QoS that was actually delivered. Menasce & Almeida (2002) discussed the basic metrics and models in Web service resource management.

Resource optimization problems in distributed systems have also been widely studied based on either trustworthiness or response time. For example, Menasce & Casalicchio (2004) used the average response time as their QoS metric in studying Grid resource management. A collection of such papers can be found in Nabrzysbi et al. (2004).

Most current studies address only one of these QoS metrics but are reluctant to deal with all of them together due to the complexity of these metrics. Our first challenge is that we need to consider a resource optimization problem subject to all three QoS metrics, expressed by trustworthiness, percentile response time and service availability. The second challenge is that we need to provide the accurate calculation of these three QoS metrics in a variety of distributed service computing models. For example, in the case of the response time as a performance metric, the average response time to process and complete a job is typically used in the literature. But, an average time may not be of real interest to a customer. For example, response time may be substantially high sometimes due to a disaster event such as a network failure. It will thus result in a high average response time even if the rest of response times are relatively low. This means that the average response time is misleading in this case. Thus, a statistically bounded metric, that is, percentile response time, is used where we take into a consideration of this abnormal situation as mentioned previously. The percentile response time is more realistic than the average response time. However, the calculation of percentile response time is often very difficult in a distributed service computing system.

OUR PROPOSED SOLUTION AND PRACTICAL IMPLICATIONS

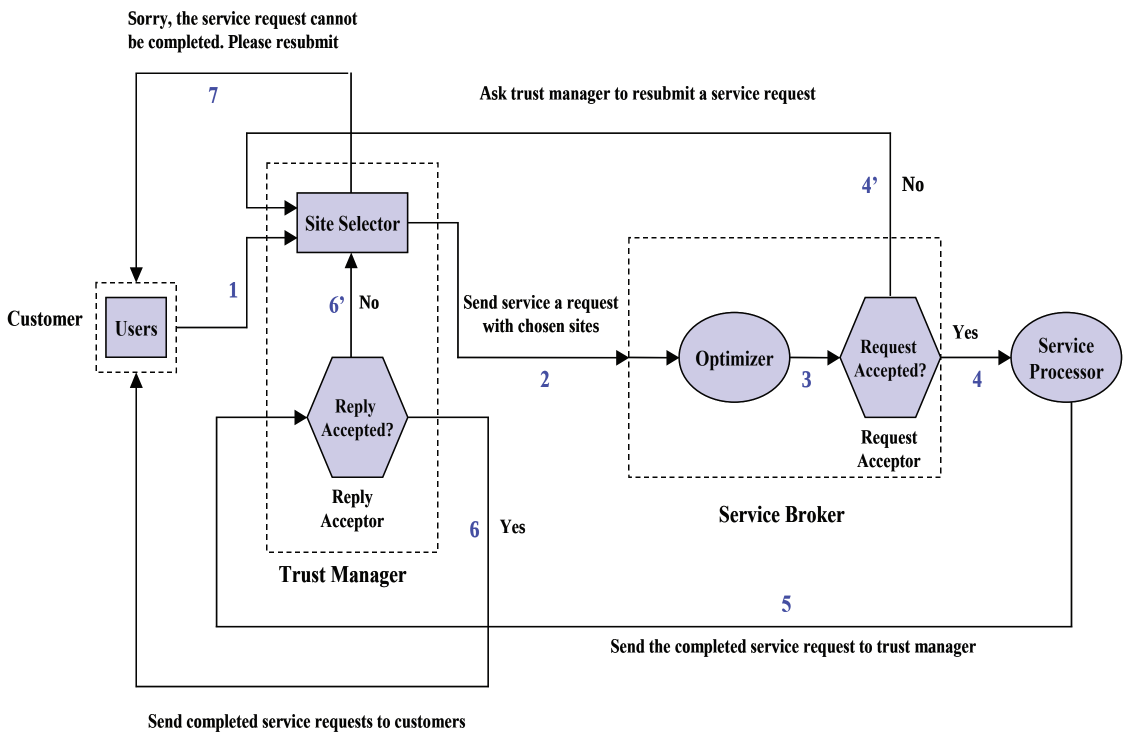
This part discusses a framework for solving the security-aware resource optimization problem in detail, our algorithm for solving the optimization problem and its practical implications.

A Framework for Solving the Security-Aware Resource Optimization Problem

Services components from several universal service providers can be flexibly integrated into a composite service with cross-language and cross-platform regardless of their location, platform, execution speed, and process. Delivering quality services to meet customer's requirement under SLA is very important and challenging due to the dynamical and unpredictable nature of Web service applications. In the part, we propose a Web service framework for solving the Security-aware resource optimization problem as shown in Figure 2. The framework consists of a customer, a trust manager, a service broker and a service processor. The functions of these service entities are explained as follows.

- The customer represents a business that negotiates a contract for particular Web services with a service broker and submits a service request to be processed by a number of resource sites.
- The customer consists of a number of business end-users (simply called users), and the service request defines service jobs generated by the users at a given rate, and QoS requirements with a fee.
- The trust manager is an entity that plays in the following roles:
    - **Service integration:** According to the customer's request, it integrates services chosen from a number of resource sites.
    - **Trust monitoring and determination:** It monitors and determines the trustworthiness of service sites.
    - **Service selection:** It chooses service sites that meet trust level requirements predefined by the customer.
    - **Service orchestration:** It defines the sequence and conditions in which one

Figure 2. A Framework for solving the security-aware resource optimization problem



Web service invokes other Web services in order to realize some useful function. That is, it defines the pattern of interactions that service components must follow in order to process the customer's requests.

Based the customer request, the trust manager generates a new service request that contains a set of service sites selected by the trust manager.

- The service broker is an entity that represents service sites. After it receives the trust manager's service request, the service broker optimizes the number of service resources in each service site to ensure the customer's QoS requirements. The service resource could be a piece of software, hardware, or both and it plays a key role in completing the customer's request. In this chapter, we consider it as a hardware device, such as a blade, a CPU, a storage device, a disk, a router and so on.
- The service processor is an entity that manages and instructs service sites to process the customer request based on the trust manager's service request. The service sites are entities that provide specific services. A single or multiple service providers may own these service sites. Each site consists of a number of service resources managed by a service manager.

Furthermore, we outline the workflow for an approach to solving the security-aware resource optimization problem in the framework as shown in Figure 2.

The workflow can be presented in the following steps.

Step 1: A customer submits a service request to the trust manager who represents the customer. Let us recall that the customer represents a business that consists of a number of users within this business, and the service request consists of a number of service jobs generated by users.

Step 2: The site selector who is part of the trust manager selects service sites with the trust indices that meet the customer's trust requirement. Then, the trust manager submits the customer request to the service broker.

Step 3: An optimizer who is part of the service broker runs an optimization algorithm to find the number of servers required at each resource site to ensure the customer's SLA guarantee. Then, it will decide whether to accept the service request based on the profitability of the service broker.

Step 4: If the service broker can make an acceptable profit with a service provider who owns these chosen service sites, then the service request is accepted. Subsequently, the service request is submitted to a service processor whereby these chosen service sites process the service request based on a rule defined in the trust manager's request. If the service request is not accepted, go to Step 4'. The service processor is an entity that manages service sites.

Step 5: After these chosen service sites finish the service request, the service processor sends the completed service request back to the trust manager. Then, the reply acceptor who is part of the trust manager will decide whether to accept the service reply based on the trust indices of these chosen sites at the time when it receives the response from the service processor.

Step 6: If the trust indices of these chosen sites meet the customer's requirement, then the service reply is accepted. Subsequently, the completed service request is forwarded to the customer.

Step 6': If the trust indices of these chosen sites do not meet the customer's requirement, then the service reply cannot be accepted. Subsequently, the service acceptor forwards the service request to the site selector and asks it to resubmit the service request.

Step 4': If the service broker cannot make an acceptable profit with a service provider who owns these chosen service sites, then the service request cannot be accepted. Thus, the service broker notifies the trust manager to resubmit the service request by re-selecting service sites.

Step 7: After the number of resubmissions surpasses a threshold predefined by the trust manager or the service broker, the trust manager will notify the customer that its request cannot be completed. Then, the customer needs to either abort the service request or modify the QoS requirement with a fee accordingly.

Remarks

- The trust manager and the service broker represent the customer and the chosen service sites respectively. Hence, we assume that they only receive service commissions from the customer and the service processor respectively based on the number of completed services. In general, the trust manager may be a business entity that not only receive a service commission from the customer, but also make money since it may work with the service broker to find service sites that meet predefined trust requirements but also the chosen service sites have an overall lower fee.
- The trust manager selects service sites on behalf of its customer. Hence, it does not care about how much cost is needed to process the customer's request. But, if the number of service resubmissions surpasses a threshold predefined by the trust manager or the service broker, then the trust manager will notify the customer that her/his request cannot be completed, as mentioned in Step 7.
- If the trust manager constantly resubmits a service request to the service broker, then the service request submissions constantly consume the service broker's resource. Thus, a denial of service attack may occur when an attacker impersonates the trust manager to keep submitting bogus service requests to the service broker. We do not analyze such an attack here since it is beyond the scope of our discussion in this chapter.
- The aforementioned steps can be regarded as an SLA negotiation process between the trust manager and the service broker. The process is often finished before service delivery. That is, the SLA is static. However, the SLA can be dynamically changed as well. In this case, the SLA will be periodically verified and/or negotiated. The length of the period of time for the verification and/or negotiation depends on individual service types. It may be a week, or a month.

An Approach for Solving the Security-Aware Resource Optimization Problem

As we saw previously, we have proposed a security-aware Web service model, and provided a Web service framework for studying the model. The framework consists of a customer, a trust manager, a service broker, and a service processor. In this framework, an optimizer within the service broker is an important key component. It calculates the number of service resources required to ensure that the response time of a service request meets the requirement of predefined percentile response time under a given fee. The security-aware resource optimization problem can be constructed by minimizing the total cost of service providers while satisfying SLA guarantees. It will be formulated as a mathematical minimization problem later.

In this part we first use a queueing network to model m service sites, formulate the security-aware optimization problem as three sub-problems. Then, we demonstrate how to develop an algorithm to solve the security-aware resource optimization problem.

Without any confusion, we reuse m (0<m£M) as the number of resource sites necessary for processing a customer's service job. Assume that the trust manager first selects m resource sites from the M ones. We only consider the case of single customer services in this chapter. Our discussion to follow can be easily extended to the case of multiple priority customer services.

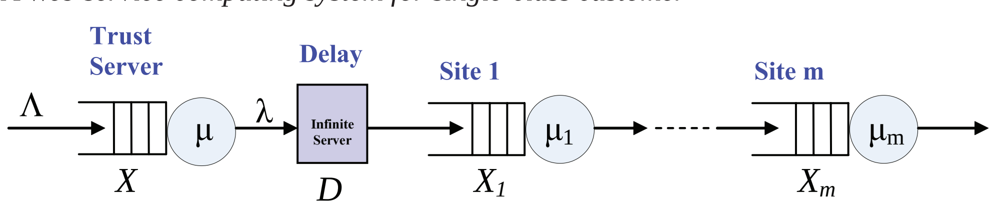
We can model those m resource sites as a queueing network, for example, the one as shown in Figure 3. Without loss of generality, we assume that the first m sites are chosen. In Figure 3, the tandem queueing network consists of a trust server and m stations numbered sequentially from 1 to m. The trust server represents the trust manager, and each station represents a resource site. Each station carries out a particular function, such as a database server, a computing server, a file server, or a Web server. The execution procedure of a service request may be complicated in the real world, but the main idea of our proposed modeling approach can be extended to describe any collaborative relationship among resource sites as long as the relationship can be quantified.

In Figure 3, each station j is modeled as a single FIFO queue served by n_j identical servers, each providing a service at the rate μ. Let λ be the external arrival rate to the infinite Server, and let λ and λ_j be the effective arrival rates to the infinite server and station j (j=1, 2, …, m). The notion of server here is defined as a service resource at each site that processes users' jobs. For example, as we mentioned early, it could be a blade, a CPU, a disk, a storage device, and so on. We assume that all service times are exponentially distributed and the external arrival to the trust server occurs in a Poisson fashion. (We relax the Poisson assumption on the external arrival from customer requests in Xiong & Perros (2008).) The trust server provides a service at the rate μ. In Figure 3, the infinite server represents the total propagation delay from the user to the service provider and back and also from station 1 to m.

In this research we are interested in minimizing the overall cost of the aforementioned Web service computing system so that the desired SLA is guaranteed. Before presenting an algorithm for solving the minimization problem, we first need to calculate the response time of a customer's service request. Recall that the response time refers to the time elapsed from the moment a customer's service job joins the computing system to the time it departs. That is, the response time of a service request is equal to a sum of the processing time of the trust manager and the execution time of chosen service sites for the customer request. It is the most important QoS metric. The response time also reflects security behavior and the availability of services in some degree. For example, an extremely high response time may imply that some service sites are under Denial of Service (DoS) attacks or unavailable in a certain period of time.

In our approach we considered the percentile response time that may better address a customer's concern as opposed to the average response time that is commonly used in the literature. Let T be a random variable that represents the response time

Figure 3. A Web service computing system for single-class customer



of a service job. Also, assume that $f_T(t)$ and $F_T(t)$ are the probability and cumulative distributions of T respectively. Denote the overall service cost of processing a service request by

$$g(n_1, n_2, \ldots, n_m) = \sum_{j=1}^{m} n_j c_j \qquad (1)$$

where $n_1, n_2, \ldots, n_m$ are the number of servers allocated to a specific stream of jobs with arrival rate λ, each server with the costs of $c_1, c_2, \ldots, c_m$.

Then, the security-aware resource optimization problem can be formulated as the following three sub-problems:

Sub-problem 1: Select m resource sites within a predefined trust index at time $t=t_k$

Sub-problem 2: Solve for n_j in the m-dimensional integer optimization problem:

$$\min_{n_1, n_2, \ldots, n_m} g(n_1, n_2, \ldots, n_m) \qquad (2)$$

Under the constraint of percentile response time:

$$F_T(t = T^D) \geq \gamma\% \qquad (3)$$

and the constraint of service availability for service sites j=1,2,…,m:

$$P_j(n_j, N_j) \geq \delta_j\% \qquad (4)$$

where T^D is a desired response time defined by a customer, γ% is a desired percentage of the response time, and $\delta_j\%$ is a desired percentage of service availability at site j. $P_j(n_j, N_j)$ is the probability that there are at least n_j servers available among N_j servers at service site j.

Sub-problem 3: Update the trust indices of all M sites based on the activity during the time interval $[t_k, t_k+T^D]$. Then, the trust manager decides if a completed job is accepted. If each trust index at those selected sites that complete the service job meets a redefined index value, then the completed job is accepted. Otherwise, then the completed job is discarded and the trust manager needs to resubmit the job.

Note that in the model the verifying time of a trust index is ignored since we are interested in the processing time of a job at resource sites. While Sub-problems 1 and 3 are solved at the customer side to ensure a customer service received from reliable resource sites, Sub-problem 2 is solved within the domain of the service broker who represents these resource sites. Hence, in the current model the trustworthiness of resource sites in Sub-problems 1 and 3 are not considered as a constraint for the optimization problem in Sub-problem 2.

To solve Sub-problems 1 and 3, we need to constantly update the trust index of each service site. We consider the discrete time $t_1, t_2, \ldots, t_k, \ldots$ in an increasing order (k=1,2,…). For each site j, its trust index at time t_k is defines as a weighted sum of its trust index at time t_{k-1} and a satisfactory ratio for services completed at time between t_{k-1} and t_k. The satisfactory ratio is equal to the number of satisfactory services assessed by customers (i.e., through customer's feedback) divided by the total number of service submissions from time t_{k-1} to t_k. Its detailed discussion can be found in Xiong & Perros (2006a).

The aforementioned resource optimization problem as described in Sub-problem 2 is to find a minimal cost presented in (2) such that γ% of the time a customer's response time is less than a predefined value T^D, and $\delta_j\%$ of time the selected resource sites have n_j servers available for the job.

Furthermore, to solve the minimization problem as described in Sub-problem 2, we need to derive the calculation of the cumulative distribution of response time T, $F_T(t= T^D)$, for inequality (3) and the probability $P_j(n_j, N_j)$ for inequality (4) respectively.

First, the complexity of the calculation for $F_T(t= T^D)$ is determined by the workload of service requests. When the workload can be modeled as the queueing network depicted in Figure 3, the

Laplace transform of the probability distribution of the total response time among m service sites is a product of each Laplace transform of such a probability distribution at each service site. Subsequently, we can derive the calculation of the probability and cumulative distributions of the total response time among m service sites. Their detailed calculation can be found in Xiong & Perros (2006b). For example, as derived in Xiong and Perros (2006b) for the case when each service site is equally utilized, the cumulative distribution of the response time T for the service model described as a queueing network depicted in Figure 3 is

$$F_T(t) = L^{-1}\left(\frac{\mu(1-\rho)}{s+\mu(1-\rho)}.\frac{\lambda}{s(s+\lambda)}.\frac{\hat{a}}{(s+\hat{a})^m}\right) \quad (5)$$

where $\hat{a} = \psi(n_j)\mu_j(1-\rho_j)$, $\rho = \frac{\lambda}{\mu}$, $\rho_j = \frac{\lambda}{\mu_j}$, $\psi(n_j) = \xi^{\log_2 n_j}$, and ξ is a server basic scaling factor from one to two servers for j=1,2,…,m. L^{-1} represents an inverse Laplace transform.

Second, a two-state Markov chain with the states "up" and "down" is used to study the service availability at each service site. Assume that each server fails at a rate of a_j, and recovers (i.e., it is put back into operation) at a rate of b_j at site j. That is, the failure rate a_j is the state of transition from "up" to "down," and the recovery rate b_j represents the rate of transition from "down" to "up." Thus, the probability that i servers are down at site j is given by

$$p_j^i = \frac{N_j!}{i!(N_j-i)!}\eta_j^i p_j^0 \quad i=1,\ldots, N_j$$

where $\eta_j = \frac{a_j}{a_j+b_j}$ is the server unavailability rate, and p_j^0 is given by

$$p_j^0 = [N_j!\sum_{i=0}^{N_j}\frac{\eta_j^i}{i!(N_j-i)!}]^{-1}$$

Thus, the probability that at least n_j servers are available at site j is:

$$P_j(n_j, N_j) = p_j^0\sum_{i=0}^{N_j-n_j}\frac{N_j\eta_j^i}{i!(N_j-i)!} \quad (6)$$

To ensure the fairness of all service sites, we can assume that each service site as depicted in Figure 3 is equally utilized. Then, Sub-problem 2 in the security-based resource optimization problem is solved using the following algorithm.

Algorithm

1. Find the minimal number of servers at each service site such that the constraint of percentile response time given in (3) is satisfied, denoted as $n_j(T)$.
2. Find the minimal number of servers at service site such that the constraint of service availability given in (4) is satisfied, denoted as $n_j(a)$.

Then, the minimal number of servers at each service site is $n_j = \max\{n_j(T), n_j(a)\}$ required to ensure that both constraints of percentile response time and service availability are guaranteed, where j=1,2,..,m.

Note. In the algorithm, Steps 1 and 2 are solved numerically and iteratively since (3) and (4) are usually complicated.

Algorithm's practical implications: The key advantage of the algorithm is that we only need to seek for the minimal number of servers required to ensure one of two constraints at each service site. The algorithm reduces the m-dimensional optimization into two one-dimensional minimization problems, which greatly reduces the computational cost of solving the security-aware optimization problem. Thus, our proposed solution can be easily realized in real-time applications.

Example

We have demonstrated how the aforementioned algorithm can be used to solve the aforementioned resource allocation problem. All simulation experiments were performed on a Dell PC running Windows XP, which has a 3.0 GHz Pentium 4 process and 1 GB memory. The queueing network as depicted in Figure 3 was simulated in Arena 10.0, a simulation software. The aforementioned algorithm was implemented in Mathematica.

Let's consider a seven-resource site example. For simplicity, all parameters to follow are unitless. We choose ξ=1.5, λ=100, T^D=0.16, γ=97.5, μ=200, δ_j=99.999, c_j=2, and N_j=200. The service rates of seven service sites are given by μ_1=18, μ_2=80, μ_3=μ_7=35, μ_4=41, μ_5=15, and μ_6=25.

We reasonably assume that each site is equally utilized to ensure the fairness of all sites. Then, by using Step 1 in the aforementioned algorithm, we can find the minimal numbers of servers for service sites 1 to 7 are 62, 5, 20, 16, 84, 35 and 20 necessary to ensure 97.5% response time guarantee. That is, n_1(T)=62, n_2(T)=5, n_3(T)=20, n_4(T)=16, n_5(T)=84, n_6(T)=35, and n_7(T)=20. Similarly, by using Step 2 in the aforementioned algorithm, we can find the minimal numbers of servers for service sites 1 to 7 are 10, 7, 22, 18, 15, 23, 10 necessary to ensure 99.999% service availability guarantee. That is, n_1(a)=10, n_2(a)=7, n_3(a)=22, n_4(a)=18, n_5(a)=15, n_6(a)=23, and n_7(a)=10. Therefore, the minimal numbers of servers for service sites 1 to 7 are 62, 7, 22, 18, 84, 35 and 20 necessary to ensure both 97.5% response time and 99.999% service availability guarantees. Thus, the total cost is:

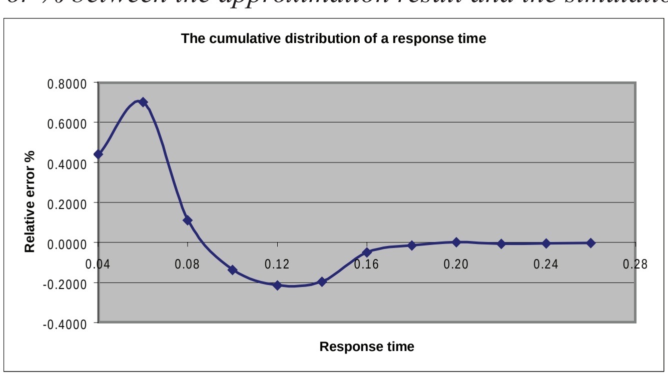
$$g(n_1, n_2, \ldots, n_7) = \sum_{j=1}^{7} n_j c_j = 496$$

When each service site is equally utilized, the accuracy of our proposed algorithm for solving the security-aware optimization problem depends on the quality of the obtained cumulative distribution given in (6). For this purpose, we simulated the aforementioned queueing network model in Arena 10.01, and compared the simulation result with the approximate result based on the formula (6). Figure 4 gives their relative error % that was calculated by

$$\text{Relative error \%} = \frac{Approximation\,\mathrm{Re}\,sult - Simulation\,\mathrm{Re}\,sult}{Simulation\,\mathrm{Re}\,sult} \times 100$$

Figure 4 validates that the approximate result is very close to the simulation one. Moreover, the exact optimal number of servers, obtained by exhaustive search using the simulation model, is consistent with the ones obtained as previously shown.

Figure 4. Relative error % between the approximation result and the simulation result



CONCLUSION AND FUTURE DIRECTIONS

We considered three QoS metrics, namely, trustworthiness, percentile response time and availability. Then, we studied all these three QoS metrics in a security-aware resource optimization problem that typically arises in Web services, and formulated the security-aware resource optimization problem as an optimization problem under SLA constraints in the case of single class. In our approach we considered the percentile response time that may better address a customer's concern as opposed to the average response time that is commonly used in the literature. By the use of an effective and accurate numerical solution for the calculation of the percentile response time (also refer to Xiong & Perros (2006b)), we solved the optimization problem using an efficient numerical procedure. Our numerical validations showed that our proposed algorithm has provided a good accuracy.

In the chapter, we provide a preliminary study of the security-aware resource optimization in distributed computing. The study raises several interesting questions. We briefly discuss some of them as follows.

We introduce a trust manager who is a trusted agent representing customers, and a service broker who a trusted agent representing service sites. The failure of a single point on either the trust manger or the service broker could be catastrophic. Hence, a decentralized mechanism should be introduced for the replacement of the centralized design. But, how much performance overhead would be added when a decentralized mechanism is in place? Additionally, if a service provider could not provide QoS as defined in an SLA, what penalty should the service provider encounter and how to implement a penalty mechanism? Furthermore, the trust manager selects service sites by using their trust indices that are partially based on the feedback of customers. Then, the question is how to deal with customer's negative feedback. Additionally, under the assumption that each site is equally utilized to ensure the fairness of these service sites, we can convert the security-aware resource optimization described as an m-dimensional optimization problem subject to two constraints into two one-dimensional optimization problems. This assumption is quite reasonable from a practical point of view. However, how to efficiently solve the security-aware optimization problem without the assumption purely from a theoretical point of view? That would also be one of the future studies.

REFERENCES

Azzedin, F., & Maheswaran, M. (2002). Evolving and managing trust in Grid computing systems. In *Proceedings of the IEEE Canadian Conference on Electrical Computing Engineering (CCECE '02).*

Brown, A., & Patterson, D. (2000). Towards Availability Benchmarks: A Case Study of Software RAID Systems. *In Proceedings of 2000 USENIX Annual Technical Conference.*

Chandra, A., Gong, W., & Shenoy, P. (2003). Dynamic Resource Allocation for Shared Data Centers Using Online Measurements. In *Proceedings of Eleventh International Conference on Quality of Service (IWQoS 2003).*

CISCO (2004). Increasing network availability, from *http://www.cico.com/wrap/public/779/largeent/learn/technologies/ina/IncreasingNetworksAvailability-Abstract.pdf.*

Dobson, G. (2002). Quality of service in service-oriented architectures, from http://digs.sourceforge.net/papers/qos.html.

Jurca, R., & Faltings, B. (2005). Reputation-based Service Level Agreements for Web services. In *Proceedings of Third International Conference on Service Oriented Computing* (ICSOC 2005), Amsterdam, The Netherlands.

Keller, A., & Ludwig, H. (2003). The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web services. *Journal of Network and Systems Management, Special Issue on ``E-Business Management*, (pp. 27-33).

Menasce, D., & Almeida, V. (2002). *Capacity planning for Web services: Metrics, Models, and Methods*. Prentice Hall PTR, Upper Saddle River, NJ.

Menasce, D., & Casalicchio, E. (2004). A Framework for Resource Allocation in Grid Computing. In *Proceedings of the 12th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'04)*, (pp. 259-267).

Levy, R., Nagarajarao, J., Pacifici, G., Spreitzer, M., Tantawi, A., & Youssef, A. (2003). Performance Management for Cluster Based Web services. In *Proceedings of the 8th IFIP/IEEE International. Symposium on Integrated Network Management (IM2003).*

Nabrzysbi, J., Schopf, J., & Weglarz, J. (2004). *Grid Resource Management*. Boston, MA: Kluwer Academic Publication.

Tosic, V., Patel, K., & Pagurek, B. (2002). WSOL - Web Service Offerings Language. *Lecture Notes In Computer Science, 2512*. Revised Papers from the International Workshop on Web services, E-Business, and the Semantic Web, (pp. 57-67).

Vu, L., Hauswirth, M., & Aberer, K. (2005). QoS-based Service Selection and Ranking with Trust and Reputation Management. *Infoscience's Technical Report*.

Wickramage, N., & Weerawarana, S. (2005). A Benchmark for Web Service Frameworks. In *Proceedings of the 2005 IEEE international conference on service computing*, (pp. 233-242).

Xiong, K., & Perros, H. (2006a). Security-aware resource allocation in Web services. In *Proceedings of the International Conference on Web Services*, Chicago, IL.

Xiong, K., & Perros, H. (2006b). Resource optimization subject to a percentile response time SLA for enterprise computing. In *Proceedings of the IEEE Globecom: Quality Reliability and Performance Modeling for Emerging Network Services (Globecom-QRPM),* San Francisco, CA.

Xiong, K., & Perros, H. (2008). SLA-based service composition in enterprise computing. In *Proceedings of International Workshop on Quality of Service*. Enschede, The Netherlands.

Zhang, Q., Yu, T., & Irwin, K. (2004). A classification scheme for trust functions in reputation-based trust management. *In Proceedings of the International Workshop on Trust, Security, and Reputation on the Semantic Web,* Hiroshima, Japan.

Ziegler, C., & Lausen, G. (2004). Spreading Activation Models for Trust Propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-commerce, and e-Service (EEE '04).*

KEY TERMS

Availability: It is the percentage of time that a service provider can offer services.

Distributed Service Computing: It is a method of computer service processing in which different parts of a service program run simultaneously on two or more computers that are communicating with each other over a network.

Quality of Service (QoS): It refers to the capability of a service provider to provide a better service to its customers. The elements of QoS often include security, performance, and availability that are defined in SLA in service computing.

Response Time: It is the time for a service request to be satisfied. That is, this is the time it takes for a service request to be executed on the service provider's multiple resource sites. Note that the definition is usually known as the *turnaround time* in a computer system. However, we have proposed a queueing network method to calculate the response time in the chapter. Thus, the definition of response time is given in the chapter from a queueing theory point of view. That is, it is the total time that a job spends in the queueing system.

Service Composition: It integrates services chosen from a number of universal service sites regarding of their location, platform and execution speed.

Service Level Agreement (SLA): It is a service contract where the level of service is formally defined through a negotiation between customers and their service provider, or between service providers.

Trust: It is defined as a firm belief in the competence of a resource site that acts as expected. "Trust" is used to deal with the notion of the trustworthiness.

Throughput: It is the service rate that a service provider can offer. It is defined as the number of service requests whose processing is completed in a single unit of time.

Web Service: It a software system designed to support an interoperable interaction between computers over a computer network.

Chapter LX
Trends, Optimization and Management of Optical Networks

Jadranka Skorin-Kapov
Stony Brook University, USA

Darko Skorin-Kapov
Adelphi University, USA

ABSTRACT

Due to the enormous proliferation of internet traffic, including video on demand and transfer of data-rich documents, the use of optical networking is mandatory, resulting with high capacity and speed as needed. In this chapter we review optical networking, its current issues, trends, and ways to efficiently design and manage a wide range of optical networks. We address the following: an overview of optical networking (including its current capabilities, virtual versus physical topology, optical node components), placement of "special equipment" nodes and components, and design of network topology. Efficient management and utilization of geographically different networks (such as wide area networks, metro networks, and access networks) present different optimization problems dealing with the placement of special optical equipment. We summarize current algorithmic approaches and suggest possibilities for further research.

OPTICAL NETWORKS: TRENDS, OPTIMIZATION AND MANAGEMENT

Introduction

The ever-increasing demands for internet usage and transfer of data-rich documents mandate the use of optical networking. Light is a form of energy represented as an electro-magnetic wave. It can be induced by stimulated emission (by lasers) producing photons that can travel in a fiber optic medium. As a wave, it has a *wavelength*, i.e. a distance between repeating units of its pattern. Currently, wavelengths are measured in 1000s of nanometers. *Wavelength Division Multiplexing (WDM)* is a technology that multiplexes different wavelengths onto a single fiber, in effect multiplexing various optical signals (channels). This increases the capacity of the optical medium. The capabilities of optical networks constantly increase, and in 2007 they are as follows: a wavelength channel can carry up to 40 Gb/s data; transmission supports up to 160 wavelengths per fiber using WDM; more than 800 fibers can be packed per cable. Just to illustrate the enormous speed, note that one DVD movie will be transferred in a second if there is a speed of 5 Gb/s. Hence, 40 Gb/s amounts to transferring about eight full DVD movies in a second!

Wavelength is inversely proportional to frequency (smaller wavelength, higher frequency) and it is related to the speed of data transfer. A measure of frequency range, measured in Hz, is called the *bandwidth*. Hence, "bandwidth" is closely related to data transmission rates. Nowadays customers require more and more bandwidth and just adding more wavelengths, without trying to bring down the unit bandwidth cost, will not be economically justified: the increase in bandwidth would not be followed by an adequate decrease in equipment cost. Hence, there is a need to carefully design and manage networks, using optimization techniques to efficiently utilize network resources.

In sum, the main current trend in telecommunications is the increased demand for high bandwidth and speed motivated by bandwidth-hungry applications such as the use of the internet, transmission of video and interactive gaming. A related trend is unpredictable bursty traffic due to heavy and unpredictable downloads. Traffic due to the transmission of voice can be modeled by a Poisson probability distribution, however, bursty traffic is self-similar and presents problems when allocating resources based on predictions; this creates design challenges when trying to allocate network resources and motivates the need for so-called "intelligent networks" which can perform automated provisioning. Another trend in contemporary telecommunications is the increased need for transparency and reconfiguration in order to achieve the required speed and to deal with un-predictable traffic. Finally, the design of network topology is moving towards irregular meshes since they are more scalable and fit better the underlying physical network of fiber cables.

As already indicated, the challenge in telecommunications planning and management for business arises from trying to consolidate revenue growth with enormous bandwidth demand. If there is no optimization of network resources, the unit bandwidth cost will be too high, resulting with too slow revenue growth, or with no growth at all. This indicates a need to design and manage network architectures with reduced unit bandwidth cost. The unit bandwidth cost may be reduced by operating networks with appropriately placed and reduced switching/routing equipment and port cards. In other words, in order to achieve savings on operating expenses (OPEX), there is a need to maximally utilize optical technology in creating a network with a small number of expensive devices, yet a network capable to deliver the required capacity and Quality of Service (QoS). This necessitates careful decisions about placing devices such as: amplifiers, passive splitters, monitors, wavelength converters, optical cross-connects, i.e. switching and routing equipment. In addition, reduced port cards call for a smaller number of transmitters (lasers) and receivers (filters). This chapter will address a number of equipment placement problems arising in optical networking, in geographically various networks such as backbone, metro, and access networks. In addition, we consider strategies for traffic grooming which are necessary to efficiently utilize the huge available bandwidth, in turn providing economic incentives for users as well as network operators.

BACKGROUND

Over the course of the last 15 years optical networking evolved, giving rise to various generations of optical networks as follows:

- *1st Generation*, Opaque networks: Every node requires electro/optic/electro (e/o/e) conversion. Only the links are made of fiber optics cables, but at every node the optical signal is converted to an electronic signal. This significantly reduces the speed of transfer. An example of opaque networks is the so-called *Broadcast and Select* network.
- *2nd Generation*, Transparent or Translucent network (*Wavelength Routed Networks*): Some nodes do not require e/o/e conversion because they are equipped with optical cross-connects. This increases the speed. The networks are circuit-switched: a lightpath (i,j) is established when a transmitter of node i and a receiver of node j are tuned to the same wavelength. A transparent network is all-optical; translucent network has nodes requiring e/o conversion, for example for monitoring purposes or regeneration.
- *3rd Generation* of optical networks will allow for *sub-wavelength traffic granularity*. Instead of circuit switched, they will be burst-switched, and further they will be packet-switched. Sub-wavelength granularity is a necessity for efficient usage of huge bandwidth. However, it seems that we are a long way from implementation of either optical burst or packet switched networks. As opposed to sub-wavelength granularity, an interesting approach is to deal with super-wavelength granularity (waveband) as a way to bundle together a number of wavelengths in order to reduce the size (number of ports) of an optical switch in wide area networks.
- *4th Generation* networks are the so-called *Intelligent Networks: Point-and-Click* with automated provisioning. The user needs to specify the origin, the destination, the required bandwidth and the required QoS, and the network's control plane and intelligent switches allocate resources and perform the request automatically, without manual intervention and the need for network reconfiguration. Relevant work includes development of Generalized Multi-Protocol Label Switch/Automatically Switched Optical Network (GMPLS/ASON) control plane. Such a plane provides dynamic and flexible bandwidth provisioning, and optimization of the use and management of network resources.

Geographical Considerations

Different geographical considerations give rise to different networks such as Wide Area Networks (WAN s) (backbone, long haul), Metropolitan networks, and Access networks. Due to varying bandwidth, bursty demand, and QoS requests, each of the networks creates its own needs for optimization with respect to design and management.

WAN (Backbone) Networks

Backbone networks connect high capacity links with traffic groomed from access and metro networks. Hence, due to a more uniform high traffic flow, *wavelength switching* and *wavelength granularity* are reasonable assumptions. Based on a given traffic matrix, the optimization problem is to decide on a number and placement of wavelength converters in order to minimize some objective, subject to a required quality of service and available capacity. Due to large distances, signal amplification is another important issue. Finally, monitoring equipment and restoration strategies have to be in place to allow secure and undisrupted transmission. Another issue relevant for WANs is *waveband switching*

(WBS). Namely, dense WDM systems (DWDM) support large number of available wavelengths (e.g. 160), hence optical cross-connects having ports for wavelength switching would have very large size. The waveband concept is introduced to indicate clustering of a number of wavelengths that are switched jointly using one switching port, hence reducing the size of an optical cross-connect. With waveband switching, a fiber carries multiple bands that are jointly demultiplexed and demultiplexing into individual wavelengths occurs only when some channels need to be added or dropped. (Cao et al., 2006) describe a framework for waveband switching, including techniques for grouping wavelengths, and show how traditional wavelength routing differs from waveband routing.

Metro Networks

Metro networks span 20-200 km. Due to various end users connecting to its ingress nodes, there is a need for traffic grooming and for sub-wavelength and wavelength granularity. The task is to simplify networking layers by new IP protocols that can forward IP packets directly over photonics, i.e. over optical switches, without the need for an intermediate data layer such as ATM.

Metro networks connect access networks with core or backbone networks. The design of an access network such as a Passive Optical Network (PON) involves the placement of passive splitters and the connection of a number of ONU (optical network units) to the OLT (optical line termination) or service provider. Hence, OLT becomes an ingress node to a transport metro network. On the other end of the geographical hierarchy, an egress metro node connects to a backbone high capacity link. It seems that the most interesting and challenging design problems appear in designing metro networks. Here is where the grooming, sub-wavelength switching based on *GMPLS* (generalized multi protocol label switching), and dynamic provisioning needs to be addressed.

In fact, metro networks can be divided into *Metro Edge* and *Metro Core*: *Metro Edge* connects to access networks while *Metro Core* connects to backbone networks. A relevant location problem is: which nodes in a metro network to equip with core capabilities? Issues relevant to *Metro Edge* network design include: sub-wavelength granularity, electronic grooming (e-grooming) versus optical grooming (o-grooming or light trails). Issues relevant to *Metro Core* network design include: wavelength granularity, placement of regenerators and converters.

Traditionally, according to Vanderhorn et al. (2006), metro networks employ SONET/SDH rings operating with fixed bandwidth defined over a rigid rate hierarchy. Recall that *Synchronous Optical Networking (SONET)* or *Synchronous digital hierarchy (SDH)* run on a physical layer, i.e. it is a method for communicating data over fiber optic cables. The basic unit of transmission for SONET is STS-1 frame (Synchronous Transport Signal-1) consisting of overhead and payload and it is transmitted in 51.840 Mb/s. The entire STS-1 frame is 810 bytes. The STS-1 frame is transmitted in exactly 125 microseconds on a fiber-optic circuit designated OC-1 (optical carrier 1). OC-N format refers to the signal in its optical form. Three OC-1 signals are multiplexed by time-division multiplexing to form the OC-3 optical circuit running at 155.52 Mb/s. Higher speed circuits are formed by aggregating multiples of slower circuits, obtaining OC-12,…, OC-768 (which operates just under 40Gb/s). Problem with SONET is inflexibility (fixed size) of optical channels. Methods to add variability are Virtual Concatenation (VCAT), Generic Framing Procedure (GFP), and Link Capacity Adjustment Scheme (LCAS).

Vanderhorn et al. (2006) report that *Metro edge rings* span about 10-40km, operate at OC-3 or OC-12 and employ add/drop multiplexers (ADMs) that connect to end users (e.g. enterprise networks or telephone public branch exchanges). The central office serves as a hub. A *Metro core* network consists of rings that interconnect major central office hubs and connect to WANs (backbone long haul networks). Metro core rings span about 40-80 km and operate on OC-48 or OC-192.

In addition to the SONET (SDH) technology, there is a next generation technology based on Ethernet, namely Ethernet over SONET/SDH (EoS) to allow for more flexibility in packet transmissions. Regardless of whether SONET/SDH or Ethernet is at a physical layer, the use of RPR (Resilient-Packet-Ring) technology (a Layer 2 standard from IEEE), based on double-ring topology and supporting fast recovery, provides enhancements adding to scalability and efficiency of metro networks. For example, RPR increases bandwidth efficiency by assigning different access priorities when sharing the network resources.

Due to sub-wavelength granularity necessitated by end users requiring partial bandwidth, metro networks employ *Multi Protocol Label Switch (MPLS.)* MPLS is a data-carrying mechanism that emulates some properties of a circuit-switched network over a packet-switched network. It lies between layer 2 (data) and layer 3(network), hence it is called "layer 2.5 protocol". Packets are forwarded along *Label Switched Paths (LSPs)* according to their labels, starting at Label Edge Router (LER) (*ingress router*), forwarding to the next router, etc. The last router in the path (*egress router*) removes the label and forwards the packet based on the header. Routers in between are called *Label Switching Routers (LSRs)*

Access Networks

Access networks are connections between end-users and metro networks. A Passive Optical Network (PON) is a point-to-multipoint network where passive (i.e. without power) optical splitters enable a single fiber to serve multiple end users. It consists of an optical line termination (OLT) (service provider) and a number of optical network units (ONUs) or optical network terminals (ONTs), and the fibers and splitters between them. The Ethernet PON (EPON or GEPON) standard was completed in 2004. It uses standard 802.3 Ethernet frames with symmetric 1 Gbs upstream and downstream rates. In 2007, its deployment was realized by Verizon, British Telecom, and others. Broadband PON (B-PON) based on wavelength multiplexing was in place in 2005, while nowadays there is Gbit PON (G-PON) with bandwidth up to 2.5 Gb/s. There is also Ethernet PON (E-PON) and the work towards 10 Gbit EPON (X-PON or 10GPON) in Japan.

Switching Granularities

Geographically different optical networks require different switching granularity in order to utilize available bandwidth more efficiently. Switching granularity can be based on circuits, packets, optical bursts, or light trails.

In a *circuit switched network* a complete path from a source to a destination is established before communications begins. When the transfer is completed, the path is removed. This is best used when there is a substantial flow between a source and a destination.

In a *packet switched transmission*, a packet consisting of load (data) and of a header is assembled and then sent from node to a node, one hop at a time. The header contains information regarding the source and destination nodes and error correcting codes. Packets are queued in buffers at intermediate nodes until sent. In electronic domain, the buffering is accomplished with random access memory (RAM). Currently, here is no equivalent of optical RAM. Short optical buffering can be accomplished by using *Fiber Delay Lines (FDLs)*.

Optical burst switching separates the data burst (payload, possibly consisting of many packets) from the control header by the interval of offset time. This is the time needed for processing of control packet in the electronic domain, in order to reserve network resources allowing the burst data to go through all-optically. The granularity of a burst is between a wavelength and a packet.

Light trail switching is a type of sub-wavelength granularity whereby intermediate nodes on a lightpath can access the wavelength channel. It is analogous to an optical bus carrying multiple connections between a source-destination pair. The light trail paradigm allows grooming in the optical domain, as opposed to electronic grooming, in effect increasing the transparency of the optical network.

MAIN FOCUS OF THE CHAPTER

The purpose of this chapter is to present a survey of optimization problems related to the design and management of Wide Area, Metro, and Access optical networks. These optimization strategies attempt to place network equipment and arrange network topologies so that the unit bandwidth cost decreases. For service providers, network cost includes the installation of links and nodes with equipment, while for network users, the cost includes the bandwidth cost and various tariffs for QoS. In this chapter we consider cost from the service provider's side and we look for strategies to achieve efficient network planning and utilization. A number of current network optimization models are presented in (Pioro & Medhi, 2004). The book discusses theory and practice in a balanced way, hence provides a very useful reference for application of various optimization models helping managers making informed decisions regarding network utilization. In the sequel, we will present an illustrative example of an optimization problem, following with a number of specific approaches for optimizing node locations in an optical network.

The basic optimization problem related to a backbone network design (assuming wavelength routed circuit switched networks) is the so-called *virtual topology/routing and wavelength assignment problem*: Given a matrix of static traffic demands between nodes and a physical topology, create a virtual topology consisting of lightpaths, route it over the physical topology, assign wavelengths, and send the traffic in order to optimize a required network performance measure (e.g. minimize congestion, or minimize the number of wavelengths used, or minimize the number of hops, or minimize the number of used add/drop multiplexers). In the dynamic case the usual objective is to minimize the blocking probability.

The combined problem gives rise to a challenging integer programming problem usually attempted with a commercial optimizer such as CPLEX for some smaller data instances. Since the problem is intractable for larger instances, often it is solved heuristically by decomposing it into a set of sub-problems such as virtual topology design, lightpath routing, wavelength assignment, and traffic allocation. As an example we present a basic formulation for the *Virtual Topology, Routing and Wavelength Assignment problem (VRWA)*, in order to illustrate a complexity of this problem, and in order to create a base for addressing placement problems presented subsequently in this chapter.

VRWA Problem: Network and Model Description

We assume that there are N nodes in the network connected via fiber links supporting up to W wavelengths (channels) by using the Wavelength Division Multiplexing (WDM) technique. Each node i has T_i transmitters and R_i receivers. The fiber links comprising physical topology are given by the NxN matrix P where $P(m,n) = 1$ if there is a physical link between nodes m and n, and $P(m,n) = 0$ otherwise. C denotes the capacity of each channel. The static traffic matrix of connection requests is given as the NxN matrix Λ where $\Lambda(s,d)$ denotes the traffic sent from source s to the destination d. The matrix is non-symmetric, i.e. the traffic from s to d can be different from the traffic from d to s.

In this setting, the task is to create a virtual topology by tuning transmitters and receivers of network nodes in order to create lightpaths through which traffic from all the sources can be sent to all the destinations. This generates different sub-problems. First, we need to establish a virtual topology described by the usable lightpaths connecting all source-destination pairs. Further, we need to provide a physical embedding of the virtual topology by establishing the actual routes over the existing fiber links and making sure that no more than the available number of wavelengths on each fiber link is used. Finally, we need to route the required traffic over the lightpaths, making sure that channel capacity is taken into account, and we need to assign wavelengths to lightpaths. A feasible solution to this complex problem needs to be evaluated in light of an objective such as minimizing cost, or maximizing QoS. In this example our objective is to minimize the maximal flow on a link, which corresponds to minimizing congestion and increasing the throughput of the network. The following variables are needed for our mathematical formulation:

- v_{ij}, $i, j = 1,\ldots,N$ is a binary variable with the following values: $v_{ij} = 1$ if there is a lightpath (a virtual link) from node i to node j ; $v_{ij} = 0$ otherwise.
- $f_{s,ij}$,$s, i, j = 1,\ldots,N$ denotes the traffic from source s sent via the virtual link (i,j).
- $p_{mn,ij}$, $m, n, i, j = 1,\ldots N$ is a binary variable with the following values: $p_{mn,ij} = 1$ if the virtual link v_{ij} uses the physical link $P(m,n)$; $p_{mn,ij} = 0$ otherwise.
- $c_{k,ij}$, $i, j = 1,\ldots N$, $k=1,\ldots,W$ is a binary variable with the following values: $c_{k,ij} = 1$ if lightpath v_{ij} uses the k-th wavelength.
- F denotes the maximal traffic flow on a virtual link (corresponding to congestion)

The formulation for minimizing the maximal flow on a lightpath (in order to allow maximal traffic scale-up), subject to constraints on virtual topology, physical embedding, wavelength assignment, and traffic routing, can then be written as follows:

$$\text{Min } F \quad (1)$$

subject to

$$\sum_{s} f_{s,ij} \le F, i \ne j \quad (2)$$

$$\sum_{j} v_{ij} \le T_i \quad \forall i \quad (3)$$

$$\sum_{i} v_{ij} \le R_j \quad \forall j \quad (4)$$

$$\sum_{i,j} p_{mn,ij} \le W \times P(m,n) \quad m \ne n \quad (5)$$

$$\sum_{n} p_{in,ij} = v_{ij} \quad i \ne j \quad (6)$$

$$\sum_{m} p_{mj,ij} = v_{ij} \quad i \ne j \quad (7)$$

$$\sum_{m} p_{ml,ij} = \sum_{n} p_{\ln,ij} \quad l \ne i \ne j \quad (8)$$

$$\sum_{i \ne l} (f_{s,il} - f_{s,li}) = \Lambda_{sl} \quad s \ne l \quad (9)$$

$$\sum_{s} f_{s,ij} \le C \times v_{ij} \quad i \ne j \quad (10)$$

$$\sum_{k} c_{k,ij} = v_{ij} \quad i \ne j \quad (11)$$

$$\sum_{i,j} p_{mn,ij}\, c_{k,ij} \le 1 \quad m \ne n \quad \forall k \quad (12)$$

The model (1) – (12) is quite general and it can be further specified by incorporating different channel capacities, and by allowing different wavelengths to be used on a lightpath. (Podnar

& Skorin-Kapov, 2002a, 2002b) considered this model on regular and arbitrary topologies and they proposed efficient heuristic algorithm based on genetic search able to improve the scalability of a benchmark data set consisting of 14 nodes in the benchmark NSFNET backbone network. In (Skorin-Kapov, 2007a) a virtual topological design problem is solved by applying rounding techniques to the linear programming relaxation of the problem. Furthermore, efficient routing and wavelength assignment using heuristics based on bin packing are proposed in (Skorin-Kapov, 2007b)

Wavelength Conversion

In the presence of optical converters, the wavelength assignment is easier and wavelengths can be re-used implying that a smaller number will be needed. The assumption of full wavelength conversion amounts to a point-to-point virtual topology. If the full conversion is not assumed, then the node placement of converters becomes a variable in the problem. Studies have shown that sparse wavelength conversion can achieve a very good performance, hence there is no need for full conversion. The formal problem dealing with converter placement can be posed as follows: "Given K converters and a network with N≥K nodes, where to place converters in order to achieve the best performance?" Factors affecting converter placement include**:** traffic at a transit node, path lengths, distances between converters, distances from the converters to other nodes, the amount of traffic that needs to be switched at a node. Heuristics for converter placement can be *topology based* or *traffic based*, and can consider either *static* or *dynamic* environment. Topology based heuristics are independent of traffic and place converters on nodes with high degree, or intermediate nodes of long paths. Traffic-load based heuristics place converters on nodes with high traffic loads and assume shortest path routing. Adaptive transit-load heuristic evaluates nodes one at a time. Greedy heuristics place one converter at a time, corresponding to the node that minimizes the blocking probability or maximizes throughput.

In the context of the *static flow matrix*, converter placement is (usually) part of the complete routing and wavelength assignment in order to maximize throughput and/or minimize cost, or minimize the number of wavelengths used in a network by deciding on the number and node placement of converters (full or limited), routing switches, amplifiers; subject to constraints on physical topology, link capacities, number of wavelengths, quality of service, etc. Some relevant work includes (Kennington at al., 2003) who formulate large mixed integer linear programs (MILPs) to solve the minimum cost routing and wavelength assignment (RWA) problems with full, sparse, and without wavelength conversion for networks with 6 – 18 nodes. The full wavelength conversion case was solvable by CPLEX, while no conversion (more challenging) and partial conversion case (the most challenging) was attempted via a tabu search heuristic. (Birkan et al. 2006) developed algorithms for routing the traffic and for designing DWDM links on networks with up to 36 nodes and 67 links. In opaque networks o/e conversion occurs in all nodes. In transparent all-optical networks, the use of optical cross-connects, amplifiers and converters allows for routing flow without o/e conversion. Cost comparison favors all-optical networks over opaque networks, with cost reductions 12% - 26%.

In the context of the *dynamic flow matrix* the objective is to minimize the blocking probability, which is the probability that a source-destination request will be blocked due to the unavailable lightpath. Blocking probability can be calculated analytically by taking into account the probabilities of wavelength usage on a link calculated from traffic loads offered to the network. Good candidate nodes for converter placement are nodes with high amount of switching traffic. (Jia et al., 2002) considered an arbitrary network topology and developed a branch & bound algorithm to solve the problem of minimizing the blocking probability subject to placing K converters. (Siregar et al., 2002) proposed a genetic algorithm to place a given number of converters for two realistic networks connecting cities in Japan. (Thiagarajan & Somani, 2003) present a problem of optimally placing a given number of converters in an arbitrary network in order to minimize the overall network blocking probability. In addition, a number of heuristic algorithms for converter placement are presented, all based on calculating appropriate weights for candidate nodes. (Jeong & Seo, 2005) present a 0/1 linear program for the placement of limited-range wavelength converters in order to a) minimize the overall network blocking probability, or b) to minimize the number of converters satisfying required performance. The blocking probability is expressed as a linear function of converter locations. Networks with 19-36 nodes were tested.

Amplifier Placement

Transporting an optical signal over some distance, or splitting it, creates the need for amplification of the signal. Optical amplifiers are devices that can simultaneously amplify optical signals using different wavelengths. The basic principle behind optical amplification is stimulated emission. Amplifiers can be used in three different places in a fiber link: (1) power amplifiers at source nodes to boost the signal before launching it; (2) line amplifiers located somewhere on the fiber link; (3) preamplifiers located at destination nodes to raise the signal prior to o/e conversion. Splitting signals creates power loss, so splitting nodes are natural candidates for power amplifiers. Types of optical amplifiers include: Raman Amplifiers, Semiconductor Optical Amplifiers (SOA), and Erbium-Doped Fiber Amplifiers.

The optimization problem can be stated as follows: Given the network topology and offered traffic, what is the minimal number of amplifiers, and where to place them in the network in order to maximize the chosen network performance or minimize network cost, subject to resource and quality of service (QoS) constraints.

Earlier works on amplifier location in broadcast-and-select networks includes (Ramamurthy et al., 1998) work on nonlinear mixed integer formulation, and (Fumagalli et al.,1998) simulated annealing heuristic. (Zhong & Ramamurthy, 2001) developed a genetic algorithm for minimizing the number of amplifiers in a wavelength-routed network assuming arbitrary permutation traffic. The placement of amplifiers has to satisfy the so-called power constraints (power levels at the start and end of each device should be within valid power level). (Deng & Subramaniam, 2003) consider amplifier placement in transparent DWDM ring networks. They consider fiber nonlinearities and various transmission impairments (e.g. propagation loss) and minimize the number of amplifiers under performance constraints in a ring network. (Tran et al., 2004) present amplifier placement algorithms for minimizing the number of amplifiers in metropolitan WDM rings, and in self-healing WDM rings. They propose nonlinear and linear programming formulations containing constraints including requirements on transmitted and received powers, the receiver's dynamic range, the amplifier gain, and the received optical signal-to-noise ratio (OSNR). The formulations are solved via commercial solvers for a ring with 6 nodes. In a related paper, (Minagar & Premaratne, 2006) also consider amplifier placement in WDM rings networks, but they add consideration of different types of amplifiers (with different gains), and the cost of such amplifiers. They extend nonlinear and linear formulations to take into account added features and to obtain solutions for ring networks with 6-10 nodes.

Regenerator Placement Problem

An amplifier (or repeater) amplifies both data and noise, so to maintain a required quality of transfer, regeneration is needed. Amplification can be done optically, while regeneration involves o/e/o transmission. The optimization problem can be stated as follows: Given a network of N nodes and a set of lightpaths L with known distances, where to place the minimal number of regenerators so that a distance between regeneration points never exceeds a required distance.

This problem is related to the placement of intelligent switches in metro networks (combination of regenerators, cross-connects), performing functions such as monitoring and grooming. (Kim et al., 2000) propose a regenerator placement algorithm in multihop all-optical networks based on dynamic programming in order to minimize the cost of placement. (Ye et al., 2003) formulate a mixed integer linear program for maximizing the number of established connections in presence of regenerators using spare transceivers in optical nodes. A heuristic based on K-least-wavelength-weight-path routing is proposed. (Gouveia et al., 2003, 2006) consider the environment as is an MPLS/WDM network. Given a WDM network topology and the offered traffic matrix with given end nodes which are *edge label switch routers (e-LSR*), the task is to decide on the number and on the placement of *core LSR (c-LSR)* and the lightpath routes that satisfy QoS requirements implemented via limited hop length. Core nodes serve as regeneration nodes and automatic bandwidth provisioning nodes. Links among core nodes correspond to label switched paths (*LSPs*) on the packet layer and they are constrained by the given number of hops, while on the optical layer there is a constraint on the maximum length WDM lightpath (to constrain optical transmission impairment). The authors present an integer linear programming formulation based on a hop-indexed approach and a two phase heuristic algorithm to solve it. The computational results include networks with up to 50 nodes.

Monitor Placement Problem

Regardless of whether one considers an access network, a metro, or a backbone network, there is a need to provide protection or backup in case of attack or failure. Hence, there is a need to place some monitoring devices to detect fault locations. Therefore, monitor placement and failure localization are related location problems.

Failure can be a result of either a *fault* or of an *attack* whereby the network stops functioning. Hardware failure (with respect to equipment and fiber) can be prevented to some extent by designing survivable networks (protection paths), whereas attacks are done on purpose and can be more difficult to manage. Different types of 'sophisticated' attacks include power jamming (amplifier attack), crosstalk attack (switch node attack), and correlated jamming (tapping attack). Failure localization is especially challenging in transparent optical networks due to the lack of o/e/o conversion over long paths, and thus monitoring must be performed in the optical domain. (In opaque or translucent networks, regenerators can serve as monitors.) Plus, optical network are vulnerable due to high traffic throughput. Hence, it is important to place monitoring devices able to send proper alarms (not sending false alarms or no alarms) when a failure occurs, and to allow localization of a failure. (Wu & Somani, 2005, 2006) consider the *crosstalk attack* and *optical power detectors* as monitoring devices. Given a network topology and a set of established lightpaths (connections), under the simplifying assumption that only one such attack occurs in the network, the authors provide necessary and sufficient conditions for detecting and localizing such an attack.

Mas et al., (2005) also consider failure location in transparent optical networks. The results suggest that proper location of monitors significantly reduces the set of potential candidates examined for a failure. Hence, restoration can be performed more efficiently. The effectiveness of the algorithm depends on the set of received alarms. This set, in turn, depends on the placement of monitoring equipment. The optimal location is the location of monitoring equipment that minimizes the number of network elements which are candidates to have a failure; in other words it minimizes the non-monitored areas.

On Access Network Design: Splitter Placement

When the access network is a passive optical network (PON), an optical line terminal (OLT) connects to optical network units (ONUs). The OLT port can support the maximum number of ONU's defined by its technology, typically 32 units. A passive splitter is a device allowing for the creation of multiple copies of the same data stream, albeit with diminished power leading to the need for amplification. If a splitter is located too close to OTL, there is no efficient utilization of fiber and bandwidth; if it is located too close to ONUs, it is difficult to have an efficient clustering of end users. Hence, it is important to place splitters at locations that would maximally utilize network resources, at the same time providing efficient data transfer.

Splitting light is easier than copying data in electronic buffer. A passive star is used in broadcast-and-select networks to create point-to-point connectivity. A passive splitter is also indispensable for multicasting. Multicast communication is supported by light trees whereby endpoints correspond to multicast users. A related location problem is to decide on which cross-connects to place power splitters. As with unicast connections employing lightpaths, a multicast connection using light trees requires solution to the routing and wavelength assignment problem. Wavelength converters relax the wavelength continuity constraint.

The *splitter placement problem* can be formulated as follows: Given a network topology and a set of multicast requests, decide on the number, capacity and placement of splitters in order to optimize a certain network characteristic.

Related work in multicast routing includes (Ali & Deogun, 2000) who consider placement of a fixed number of splitters in a WDM network without converters. They propose a two phase algorithm: first, a genetic algorithm solves the multicast RWA problem without splitting capability in order to establish the maximum number of multicast sessions; then K splitters are allocated by using either a simple heuristic based on locating splitters at nodes used as branching nodes the most, or a simulated annealing heuristic. Based on numerical experiments, the authors conclude that simulated annealing woks better, and that not more than 50% of the nodes need to be equipped with splitters in order to achieve good network performance.

Yang & Liao (2003) formulate a complex integer linear program for the design of a light-tree based logical topology in WDM networks using splitters and wavelength converters. The number of available splitters and converters is given, and the objective is to place them in order to minimize the number of wavelengths used by all light trees, subject to satisfying end-to-end delay bounds. The results are given for the 14dim NSFNET topology. (Lin & Wang, 2005) consider splitter placement in all-optical WDM networks without wavelength conversion. The objective is to locate a given number of splitters so that the average per link wavelength resource usage of multicast connections is minimized. Their approach is based on the notion of a light forest: a multicast light tree is decomposed in a group of light trees each using a different wavelength; this group is called light forest. Without construction of a light forest, each branching node should be a splitter. With light forests constructed for multicast connections, it is possible to do it without splitters at all, provided enough wavelengths. The problem is if there is an upper bound on the available number of wavelengths that can be used on a link. In such a case splitters have to be used to reduce the needed number of wavelengths. (Yu & Cao, 2006) also consider all-optical WDM networks, but release the wavelength continuity constraint. They look at a combined splitter placement and wavelength converter placement problem in the presence of splitting and conversion capacity constraints, and devise methods for both static and dynamic traffic demands. For static demands and a given a number of splitters and converters, they formulate an integer linear program combining splitter and converter placement and routing and wavelength assignment, with the objective to minimize the number of wavelengths used. The experiments are performed on the 14-dim NSFNET topology, and show that (1) both splitters and converters tend to be placed in same switches; (2) adding more splitters (converters) does not change the placements from smaller number cases. The consequence is that placing more splitters and/or converters can be done without major relocations of already placed equipment. For dynamic traffic, (Yu & Cao, 2006) propose a heuristic to sequentially place first splitters, than converters, in order to minimize the blocking probability. Heuristics using a bin packing approach for the routing and wavelength of light trees, assuming splitters at all nodes, are proposed in (Skorin-Kapov, 2006).

Interfaces Between Access and Metro Networks: Clustering and Hub Location

Considering access networks, the placement problem is to decide on the gateway nodes becoming the edge nodes of metro networks. Such nodes have to be equipped with grooming capability for combining low-speed traffic streams onto high-speed wavelengths in order to efficiently and economically use the available bandwidth. (Dutta & Rouskas, 2002) provide a comprehensive survey of traffic grooming. (Chen et al., 2006) consider clustering methods for hierarchical traffic grooming in large-scale mesh WDM networks. The network is decomposed into clusters and a hub is selected in each cluster. The hub node can groom the traffic originating and terminating locally. The problem is to determine the size and composition of clusters in a hierarchical fashion as to provide good grooming solutions. It is an integer linear program, intractable for larger instances, hence heuristics are needed.

Jackson, et al. (2007) consider a 'clustering' type problem arising in the control and management of packet-switched networks. A large capacity of a backbone transmission link (in Gb/s) allows for carrying many independent traffic flows. Each traffic flow is characterized by some parameters, e.g. QoS, bandwidth, burst size. The service provider may prefer to group similar requests into a single service level, such that any user receives at least the requested amount resource. This problem is equivalent to a *directional p-median problem in multiple dimensions*: Locate *p* supply points in order to minimize the total distance of the demand points to their nearest supply point, where the distance measure used is *directional rectilinear distance* (a supply point for a given demand point must have a set of coordinates bigger than the demand point; i.e. in a plane it must lie above and to the right of it).

Podnar, et al. (2002) propose network cost minimization using threshold-based discounting. The strategy is to offer incentives to network users to combine flows. To enjoy the discount, flows must reach a given threshold. In their model there are no hubs, but requests are grouped based on economic incentives. A generalization of this strategy can apply to traffic grooming when connection requests include multiple pairs of acceptable (QoS_i, $cost_i$) pairs. The challenge might be to perform grooming and routing so that the overall cost is minimized, subject to achieving acceptable performance. In such a scenario, a user might be inclined to accept lower QoS if that involves significant savings.

FUTURE TRENDS AND POSSIBILITIES FOR FURTHER RESEARCH

Possibilities for further research include modifications of the Routing and Wavelength Assignment problem in the presence of sub- and super-wavelength granularity (light trails, wavebands). Another avenue of research is the design of Metro Core networks by deciding on the placement of core switches: a superposition of regenerator/core network on optical layer network which, in turn, is superimposed on a physical fiber network in the presence of various constraints. Requests to utilize more efficiently network resources (e.g. bandwidth, equipment) will lead to hub-and–spoke like algorithms for clustering or grouping clients. With respect to static versus dynamic requests the issue is how to place network equipment so that dynamic requests can be accommodated without the need to re-arrange the network. For example, light trails present a generalization of circuit switching with added flexibility for a request to enter it or leave it without changing the logical topology of lightpaths.

CONCLUSION

Since optimization problems in optical networking are very complex due to multiple layers (physical, virtual WDM paths, virtual label switched paths), due to issues related to traffic grooming, wavelength usage, optical and electronic switching, and amplification and splitting, it seems that the following trend will continue: formulate a large nonlinear or linear integer program for a given set of networking assumptions, solve it for a small size with a commercial optimizer such as CPLEX, then devise a heuristic strategy to tackle partial problems or bigger sized problems. With optimization pertaining to network equipment and utilization, the planning and management of optical networking can become more efficient, resulting with smaller bandwidth unit cost, in turn leading to higher profit margins for service providers.

REFERENCES

Ali, M., & Deogun, J. (2000). Allocation of splitting nodes in all-optical wavelength-routed networks. *Photonic Network Communications, 2*(3), 247-265.

Birkan, G., Kennington, J., Olinick, E., Ortynski, A., & Spiride, G. (2006). Practical integrated design strategies for opaque and all-optical networks: optimization models and solution procedures. *Telecommuncation Systems , 31*(1), 61-83.

Cao, X., Annand, V., & Qiao, C. (2006). Framework for waveband switching in multigranular optical networks: part I - multigranular cross-connect architectures. *Journal of Optical Networking, 5*(12), 1043-1055.

Chen, B., Rouskas, G., & Dutta, R. (2006). Clustering methods for hierarchical traffic grooming in large scale mesh WDM networks. *Proceedings of the 5th International Symposium on Communication Systems, Networks And Digital Signal Processing (CSNDSP)* .

Deng, T., & Subramaniam, S. (2003). Amplifier placement in transparent DWDM ring networks. In A. Somani, & Z. Zhang (Ed.), *Proceedings of SPIE*, *5285*, (pp. 246-257).

Dutta, R., & Rouskas, N. (2002, November/December). Traffic grooming in WDM networks: past and future. *IEEE Network* (pp. 46-56).

Fumagalli, A., Balestra, G., & Valcarenghi, L. (1998). Optimal amplifier placement in multi-wavelength optical network based on simulated annealing. *SPIE International Symposium on Voice, Video, and Data Communicatios , 3531*, 268-279.

Gouveia, L., Patricio, P., & de Sousa, A. (2006). Hop-constrained node survivable network design: an application to MPLS over WDM. *Procedings of 8th INFORMS Teleommunication Conference.* Dallas, Texas.

Gouveia, L., Patricio, P., de Sousa, A., & Valadas, R. (2003). MPLS over WDM network design with packet level QoS constraints based on ILP models. *Proceedings of IEEE INFOCOM*.

Jackson, L., Rouskas, G., & Stallman, M. (2007). The directional p-median problem: definition, complexity, and algorithms. *European Journal of Operational Reseach, 179*, 1097-1108.

Jeong, H.-Y., & Seo, S.-W. (2005). A binary (0-1) linear program formulation for the placement of limited-range wavelength converters in wavelength-routed WDM networks. *Journal of Lihtwave Technology, 23*(10), 3076-3086.

Jia, X., Hu, X.-D., & Du, D.-Z. (2002). *Multiwavelength optical networks.* Kluwer Academic Publishers.

Kennington, J., Olinick, E., Ortyinski, A., & Spiride, G. (2003). Wavelength routing and assignment in a survivable WDM mesh network. *Operations Research, 51*(1), 67-79.

Kim, S.-W., Seo, S.-S., & Kim, S. (2000). Regenerator placement algorithms for connection establishment in all-optical networks. *GLOBECOM - Global Telecommunications Conference IEEE*, (pp. 1205-1209).

Lin, H.-C., & Wang, S.-W. (2005). Splitter placement in all-optical WDM networks. *Proceedings of the IEEE GLOBECOM 2005.*

Mas, C., Tomkos, I., & Tonguz, O. (2005). Failure location algorithm for transparent optical networks. *IEEE Journal on Selected Areas in Communications, 23*(8), 1508-1519.

Minagar, A., & Premaratne, M. (2006). Cost optimal allocation of amplifiers and DCMs in WDM ring networks. *Optic Express, 14*(22), 10278-10291.

Pioro, M., & Medhi, D. (2004). *Routing, Flow, and Capacity Design in Communication and Computer Networks.* The Morgan Kaufmann Series in Networking.

Podnar, H., & Skorin-Kapov, J. (2002b). A genetic algorithm for optimizing throughput in non-broadcast WDM optical networks. In A. Abraham, & M. Koeppen (Eds.), *Hybrid Information Systems* (pp. 323-338). Heidelberg: Physica Verlag.

Podnar, H., & Skorin-Kapov, J. (2002a). An application of a genetic algorithm for throughput optimization in non-broadcast WDM optical networks with regular topologies. *Mathematical Communications, 7,* 45-49.

Podnar, H., Skorin-Kapov, J., & Skorin-Kapov, D. (2002). Network cost minimization using threshold-based discounting. *European Journal of Operational Research, 137*, 371-386.

Ramamurthy, B., Iness, J., & Mukherjee, B. (1998). Optimal amplifier placements in a multiwavelength optical LAN/MAN: the unequally powered wavelength case. *IEEE/ACM Transactions on Networking, 6*(6), 757-767.

Siregar, J., Tagaki, H., & Zhang, Y. (2002). Optimal Wavelength Converter Placement in Optical Networks by Genetic Algorithm. *IEICE Trans. Commun., E86*(B(6)), 1075-1082.

Skorin-Kapov, N. (2006). Multicast routing and wavelength assignment in WDM networks: a bin packing approach. *Journal of Optical Networking, 5*(4), 266-279.

Skorin-Kapov, N. (2007a). A new objective criterion and rounding techniques for determining virtual topologies in optical networks. *IEEE Communication Letters, 11*(6), 540-542.

Skorin-Kapov, N. (2007b). Routing and wavelength assignment in optical networks using bin packing based algorithms. *European Journal of Operational Research, 177*(2), 1167-1179.

Thiagarajan, S., & Somani, A. (2003). Optimal wavelength converter placement in arbitrary topology wavelength-routed networks. *Computer Communications, 26*, 975-985.

Tran, A., Tucker, R., & Boland, N. (2004). Amplifier placement methods for metropolitan WDM ring networks. *Journal of Lightwave Technology, 22*(11), 2509-2522.

Vanderhorn, N., Balasubramanian, S., Mina, M., Weber, R. J., & Somani, A. K. (2006, March). Light-Trail Testbed for Metro Optical Networks. *Proc. of the 2nd International IEEE/Create-Net Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, TRIDENTCOM* .

Wu, T., & Somani, A. (2006). Attack monitoring and localization in all-optical networks. *Cluster Comput, 9*, 465-473.

Wu, T., & Somani, A. (2005). Cross-talk attack monitoring and localization in all-optical networks. *IEEE/ACM Transactions on Networking, 13*(6), 1390-1401.

Yang, D.-N., & Liao, W. (2003). Design of light-tree based logical topologies for multicast streams in wavelength routed optical networks. *Proceedings of IEEE INFOCOM*.

Ye, Y., Chai, T., Cheng, T., & Lu, C. (2003). Agorithms for the design of WDM translucent optical networks. *Optics Express, 11*(22), 2917-2926.

Yu, O., & Cao, Y. (2006). Placement of light splitters and wavelength converters for efficient multicast in all-optical WDM networks. *IEICE Trans. Inf. & Syst., E89-D*(2), 709-717.

Zhong, L., & Ramamurthy, B. (2001). Optimization of amplifier placements in switch-based optical networks. *IEEE International Conference on Communication,* 224-228.

KEY TERMS

Asynchronous Transfer Mode (ATM): Is a cell relay, packet switching network and data link layer protocol.

Light Amplification by Stimulated Emission of Radiation (Laser): Is a light source and converts electrons to photons, or performs electro/optic conversion.

Optical Add-Drop Multiplexers: Selectively remove a wavelength from a set of wavelengths then add the same wavelength with a different data (can add and drop traffic).

Optical Amplifiers: Amplify signals in the optical domain (regeneration); repeaters amplify after conversion to the electric domain. A single amplifier can simultaneously amplify all wavelengths. Basic types of amplifiers: Erbium-doped fiber amplifiers (EDFAs), semiconductor optical amplifiers (SOAs) and Raman amplifiers. An EDFA is a fiber segment, a few meter long. Raman amplifier is based on the nonlinear optical effect. While EDFAs provide a large gain over a short distance, the Raman amplifier provides a small gain over a large distance (and requires long fibers with high-power pump lasers).

Optical Network Components: Active components are electronically powered (lasers, wavelength shifters (converters), modulators). ***Passive*** components are not electrically powered and cannot generate light on their own (fibers, multiplexers, demultiplexers, couplers, isolators, attenuators, circulators). Optical modules are devices with active and/or passive optical elements such as transceivers, erbium-doped amplifiers, optical switches, add/drop multiplexers.

Optical Switches: Perform cross-connecting channels and provide reconfiguration either at the wavelength level, or at the fiber level.

Wavelength Converters: Convert one wavelength signal to another. SOAs (semiconductor amplifiers) are used as wavelength converter devices.

Wavelength Routed Networks: Allow all-optical connections (lightpaths) between pairs of nodes. Namely, there is no opto/electric conversion at transitive nodes; the whole path, possibly traversing many physical links, is completely in the optical domain.

Compilation of References

3GPP, at www.3gpp.org

Aarnio, A., Enkenberg, A., Heikkilä, J., & Hirvola, S. (2002) Adoption and Use of Mobile Serivces - Empirical Evidence from a Finnish Survey. *Proceeding of the 35th Hawaii International Conference on System Sciences 2002* (1454-1463).

Abdolreza E., Dominique H., & Heikki T. (2007). Determinants of customer loyalty in the wireless telecommunications industry. *Telecommunications Policy, 31*(2), 93-106.

Abdou, K., & Kliche, S. (2004). The strategic alliances between the American and German companies. A cultural perspective. *European Business Review, 16*(1), 8-27.

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review, 16*(3), 586-612.

Abrahamson, E. (1996). Management fashion. *Academy of Management Review, 21*(1), 254-285.

Acharya, J., & Yates, R. D. (2007). A framework for dynamic spectrum sharing between cognitive radios. *IEEE International Conference on Communications,1*, 5166-5171.

ACMA. (2007). *New restricted access arrangements: Restricted Access System Declaration 2007.* Retrieved 11 January, 2008, from http://www.acma.gov.au/WEB/STANDARD/pc=PC_310905

ACMA. (2007). *New rules for age-restricted internet and mobile content* (ACMA Media Release 151/2007 - 21 December). Retrieved 11 January, 2008, from http://www.acma.gov.au/WEB/STANDARD/pc=PC_310907

ACMA. (2006). *Telecommunications Performance Report 2004-05.* Sydney: Australian Communications and Media Authority (ACMA)

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in Large and Small Firms: An Empirical Analysis. *The American Economic Review, 78*(4), 678-690.

Acs, Z. J., Audretsch, D. B., Feldman, M. P. (1994). R&D Spillovers and Recipient Firm Size. *Review of Economics and Statistics, 76*(2), 336-340.

Adjih, C., Boudjit, S., Jacquet, P., Laouiti, A., Muhlethaler, P., & Mase, K. (2005). *Address auto-configuration in optimized link state routing protocol.* IETF draft, July 2005, draft-laouiti-manet-olsr-address-autoconf-01.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-laouiti-manet-olsr-address-autoconf-01.

Adner, R. (2002). When are Technologies Disruptive? A Demand-Based View of the Emergence of Competition. *Strategic Management Journal, 23*, 667-688.

Adner, R., & Levinthal, D. (2001). Demand Heterogeneity and Technology Evolution:Implications for Product and Process Innovation. *Management Science, 47*(5), 611 –628.

AePONA Causeway, at www.aepona.com

AfrISPA (African Internet Service Providers Association) (2002). *The Halfway Proposition*, Retrieved Feb. 15, 2004, from http://www.afrispa.org/initiatives.htm

Afuah, A. N., & Bahram, N. (1995). The hypercube of innovation. *Research Policy, 24*(1), 51-76.

Afuah, A., & Tucchi, C. L. (2006). *Internet Business Models and Strategies: Text and Cases*. McGraw Hill.

Aggrawal, V., Feldmann, A., and Schneideler, C. Can ISPs and P2P users cooperate for improved performance? ACM SIGCOMM computer communications review 37/3 (2007) 31-40

AgLeader Technologies (2008). http://www.agleader.com/products.php?Product=pfadvantage. (Verified January 23, 2008)

Agrawal, G. (2002). *Fiber-optic communication systems*. New York: Wiley.

Agustín, A., Muñoz, O., & Vidal, J. (2004). A game theoretic approach for cooperative MIMO schemes with cellular reuse of the relay slot. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 4*, 581-584.

Aharoni, Y. (2000). The role of reputation in global professional business services. In Y. Aharoni & L. Nachum (Eds.), *Globalization of Services: Some implications for theory and practice*. London: Routledge.

Ahlstrorn, D., & Bruton, G. (2001). Learning from successful local private firms in China: Establishing legitimacy. *Academy of Management Executive*, *15*(4), 72.

Ahmad, I., Wei, X., Sun, Y., & Zhang, Y. Q. (2005). Video transcoding: An overview of various techniques and research issues. *IEEE Transactions on Multimedia, 7*(5), 793-804.

Ahmed, P. K. (1998). Culture and climate for innovation, *European Journal of Innovation Management*, *1*(1), 30–43

Ahmed, P. K., Lim, K. K., & Zairi, M. (1999). Measurement practice for knowledge management. *Journal of Workplace Learning: Employee Counseling Today, 11*(8), 304-311.

Ahren, S., Mark, L., & Adrian, P. (2007). Efficient mechanisms to provide convoy member and vehicle sequence authentication in VANETs. In B. Crispo (Ed.), *The Third International Conference on Security and Privacy in Communication Networks*. IEEE: France.

Ai, C., & Sappington, D. E. M. (2002). The impact of state incentive regulation on the U.S. telecommunications industry. *Journal of Regulatory Economics*, *22*(2), 133-159.

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckmann (Ed.), *Action Control: From Cognition to Behavior*. New York: Springer.

Ajzen, I. (1991). The theory of planned behavior. *Organization Behavior and Human Decision Processes*, *50*, 179-211.

Aksu, A., A., & Ozdemir, B. (2005). Individual learning and organization culture in learning organizations. *Managerial Auditing Journal*, *20*(4), 422-441.

Akujuobi, C. M., & Ampah, N. K. (2007). Enterprise network intrusion detection and prevention system (ENIDPS), *Society of Photographic Instrumentation Engineers Defense and Security Symposium*, *6538*, Orlando, Florida.

Akujuobi, C. M., Ampah, N. K., & Sadiku, M. N. O. (2007). An Intrusion Detection Technique Based on Change in Hurst Parameter with Application to Network Security, *International Journal of Computer Science and Network Security*, *7, I(5)*, 55–64.

Akyildiz, I. F., Mcnair, J., Joseph, S. M., & Wang, W. (1999). Mobility Management in Next-Generation Wireless Systems. *Proceedings of the IEEE, 87*(8), 1348-1384.

Alahuhta, P., Ahola, J., & Hakala, H. (2005). *Mobilising business applications: a survey about the opportunities and challenges of mobile business applications and services in Finland*. Helsinki: Tekes.

Alam, M., & Prasad, N. (2008). Convergence transforms digital home: techno-economic impact. *Wireless Personal Communications*, *44*(1), 75-93.

Ala-Siuru, P., & Rantakokko, T. (2006). Understanding and recognizing usage situations using context data available in mobile phones. *2nd International Workshop on Personalized Context Modeling and Management for UbiComp Applications*. Berlin: Springer-Verlag.

Alcatel 8601 Parlay/OSA Gateway, at: www.alcatel.com

Ali, M., & Deogun, J. (2000). Allocation of splitting nodes in all-optical wavelength-routed networks. *Photonic Network Communications, 2*(3), 247-265.

Allan, D., Bragg, N., McGuire, A., & Reid, A. (2006). Ethernet as carrier transport infrastructure. *IEEE Communications Magazine*, *44*(2), 134–140.

Allee, V. (2001). *12 principles of knowledge management.* American Society for Training and Development (ASTD).

Alleman, J. (2002). *A New View of Telecommunications Economics.* Telecommunications Policy.

Almes, G., Kalidindi, S., & Zekauskas, M. (1999). *A One-way Delay Metric for IPPM*, IETF, RFC 2679.

Almes, G., Kalidindi, S., & Zekauskas, M. (1999). *A One-way Packet Loss Metric for IPPM*, IETF, RFC 2680.

Alpcan, T., Basar, T., & Dey, S. (2006). A power control game based on outage probabilities for multicell wireless data networks. *IEEE Transactions on Wireless Communications*, *5*(4), 890-899.

Altman, E., & Altman, Z. (2003). S-modular games and power control in wireless networks. *IEEE Transactions on Automatic Control, 48*(5), 839-842.

Altman, E., & Basar, T. (1998). Multiuser rate-based flow control. *IEEE Transactions on Communications, 46*(7), 940-949.

Altman, E., Borkar, V. S., Kherani, A. A. (2004). Optimal random access in networks with two-way traffic. *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, *1*, 609-613.

Altman, E., Boulogne, T., El-Azouzi, R., Jiménez, T., & Wynter, L. (2006). A survey on networking games in telecommunications. *Elsevier Computers & Operations Research*, *33*(2), 286-311.

Altman, Z., Skehill, R., Barco, R., Moltsen, L., Brennan, R., Samhat, A., Khanafer, R., Dubreil, H., Barry, M., & Solana, B. (2006). The Celtic Gandalf Framework. *IEEE Mediterranean Electrotechnical Conference*, (pp.5 95-598).

Amaldi, E., Capone, A., & Malucelli, F. (2003). Optimization models and algorithms for downlink UMTS radio planning. *IEEE Wireless Communications and Networking Conference*, *2*, 827-831.

Amaldi, E., Capone, A., & Malucelli, F. (2003). Planning UMTS Base Station Location: Optimization Models with Power Control and Algorithms. *IEEE Transactions on Wireless Communications*, *2*(5), 939-952.

Amanullah, M. T. O., Kalam, A., & Zayegh, H. (2005). Network security vulnerabilities in SCADA and EMS, *Transmission and Distribution Conference and Exhibition: Asia and Pasific*, (pp. 1-6). Dalian.

Amendola, G. B., & Pupillo, L. M. (2007, September). *The Economics of Next Generation Access Networks and Regulatory Governance in Europe: One Size Does not Fit All.* Paper presented at the 18th ITS Regional Conference Istanbul, Turkey.

American Bar Association (editors). (2005). *Telecom Antitrust Handbook (American Bar Association Section of Antitrust Law Monograph).*

American Productivity and Quality Centre (APQC) (2001). *Successfully implementing knowledge management.* APQC: Consortium Learning Forum Best Practice Report.

Amit, R., & Zott, C. (2001). Value Creation in E-Business. *Strategic Management Journal, 22*(6/7), 493-520.

Amoussou, G. E., Pesant, G., & Pierre, S. (2001). Affectation de cellules des commutateurs par programmation par contraintes. *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2*, 1017-1022.

AMTA. (2005). *AMTA Response to regulation taskforece isues paper 2005.* Australian Capital Territory: ACT: Australian Mobile Telecommunications Association.

AMTA. (2006). *Australian Mobile Telecommunications Association (AMTA) 2006 Annual Report*. Australian Capital Territory (ACT): Australian Mobile Telecommunications Association.

Anand, J., Rottman, J. W., & Lacity, M. C. (2006). A Review of the Predictors,Linkages and Biases in IT Innovation Adoption Research. *Journal of Information Technology, 21*(1-23).

Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., & Werthimer, D. (2002). SETI@home: An Experiment in Public-Resource Computing. *Communications of the ACM*, *45*(11), 56-61.

Anderson, E., Breitenbach, S., Burd, T., Chidambaram, N., Houle, P., Newsome, D., et al. (2001). *VoiceXML Early Adopter* (1st ed.): Peer Information Inc.

Anderson, H. R., & Mcgeehan (1994). Optimizing microcell base station locations using simulated annealing techniques. *IEEE Vehicular Technology Conference*, (pp. 858-862).

Anderson, J. C., & Narus, J. A. (2004). *Business market management – understanding, creating and delivering value*. Upper Saddle River, NJ: Pearson Education, Inc.

Anderson, S. W. (1995). A framework for assessing cost management system changes: The case of activity based costing implementation at General Motors, 1986-1993. *Journal of Management Accounting Research*, *7*, Fall, 1-51.

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S. (1987). MUNIN: A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the International Joint Conference on Artificial Intelligence*. Milan, Italy, (pp. 366-372).

Andreou, A. S. C., Chrysostomou, C., Leonidou, S., Mavromoustakos, A., Pitsillides, G., Samaras, C., Schizas (2002). Mobile commerce applications and services: A design and development approach. *The First International Conference on Mobile Business M-Business 2002*, July 8-9, Athens, Greece.

Angelou, G. N., & Economides, A. A. (2007). A decision analysis framework for prioritizing a portfolio of ICT infrastructure projects. *IEEE Transactions on Engineering Management*. Forthcoming.

Angelou, G. N., & Economides, A. A. (2007). A Real Options approach for prioritizing ICT business alternatives: A case study from Broadband Technology business field. *Journal of the Operational Research Society*. Forthcoming.

Angelou, G. N., & Economides, A. A. (2007). E-learning investment risk management. *Information Resource Management Journal*, 20(4), 80-104, October-December.

Angelou G., & Economides A. (2005). Flexible ICT investements analysis using real options. *International Journal of Technology, Policy and Management*, *5*(2),146–166.

Anne, M., Pal, S., Leif, N., & Oivind, K. (2007). On the distribution of revocation information in ad hoc networks. In P. Dini (Ed.), *The Third International Conference on Wireless and Mobile Communications* (p. 84). IEEE: USA.

Antelope Consulting (2001). *The Costs of Internet Access in Developing Countries.*

Antonelli, C. (1989). The Diffusion of Information Technology and the Demand for Telecommunication Services. *Telecommunications Policy, 13*(3), 255-264.

Antonopoulos, M. (2006). *Best practices for securing your IP address management infrastructure*. Special Report, Lucent Technologies. Retrieved November 20, 2006, from http://www.networkworldpartners.com/lucent/pdf/IPAMSRFinal_082906.pdf

APEC (2005). *Flow-based Internet Traffic Measurement and Analysis Project.* Final Project Evaluation Report, APEC #205-TC-01.5

Arbore, A., & Ordanini A. (2006). Broadband divide among SMEs: The role of size, location and outsourcing strategies. *International Small Business Journal*, *24*, 83-99.

Arduini, D., Nascia, L., & Zanfei, A. (2007). A sectorial approach to the diffusion of ICT: Empirical evidence on Italian firms. Working paper presented at the DRUID Summer Conference 2007.

Argyris, C., & Kaplan, R. S. (1994). Implementing new knowledge: The case of Activity-Based Costing. *Accounting Horizons, 8*(3), September, 83-105.

Armitage, G. (2000). MPLS: the magic behind the myths. *IEEE Communications Magazine*, *38*(1), 124–131.

Arrow, K. (1962). Economic welfare and the allocation of resources for inventions. In R. Nelson (Ed.), The Rate and Direction of Inventive Activity (pp. 609-625). Princeton, N.J: Princeton University Press.

Aschemann, G., Mohr, T., & Ruppert, M. (1999). Integration of SNMP into a CORBA and Web-based Management Environment. In Steinmetz (Ed.), *KiVS'99* (pp. 210-221).

Ashbrook, D., Lyons, K., & Clawson, J. (2006). Capturing Experiences Anytime, Anywhere. *Pervasive Computing*, *5*(2), 8-9.

Atakan, B., & Akan, O. B. (2007). BIOlogically-Inspired Spectrum Sharing in Cognitive Radio Networks. *IEEE Wireless Communications and Networking Conference, 1*, 43-48.

Atkin, D. (2002). Convergence across media. In C. Lin & D. Atkin (Eds.), *Communication technology and society* (pp. 23-42). Cresskill, NJ: Hampton Press.

Atkins, J. (2004). New tools for new networks. *Telecommunications International, 38*(8), 16-20.

Atkinson, R. (Ed.). (2004). IAB Concerns and Recommendations Regarding Internet Research and Evolution, IETF, RFC 3869.

Atkinson, R., & Floyd, S. (2004). IAB Concerns & Recommendations Regarding Internet Research & Evolution, IETF, RFC 3869.

Attewell, P. (1992). Technology Diffusion and Organizational Learning: The Case of Business Computing. *Organization Science*, *3*(1), 1-19.

Atzori, L., & Lobina, M. L. (2004). Speech *playout* buffering based on a simplified version of the ITU-T E-Model. *IEEE Signal Processing Letter, 11*(3), 382-385.

Atzori, L., Lobina M. L., & Corona, M. (2006). *Playout* Buffering of Speech Packets based on a Quality Maximization Approach. *IEEE Transaction on Multimedia 8*(2), 420-426.

Au, W., Chan, C. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*, 532–545.

Aurrecoechea, C., Campbell, A., & Hauw, L. (1998). A survey of QoS architectures. *Multimedia Systems*, *6*(3), 138-151.

Australia (2002). *Internet Traffic Exchange and the Development of End-to-End International Telecommunications Competition-Comments on the OECD Research Paper.* ITU-T Study Group 3, Delayed Contribution 22, COM 3-D 22-E, May.

Australian Bureau of Statistics, Internet activity, Latest issue (2007) <www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/Main+Features1Mar%202007?OpenDocument>

Ausubel, L. M., Cramton, P., McAfee, P., & McMillan, J. (1997). Synergies in wireless telephony: Evidence from the broadband PCS auctions. *Journal of Economics and Management Strategy*, 6(3), 497-527.

AVAYA. (2005). Migrating to Converged Networks and IP Telephony Applications:Implications for Management Tools and Managed Services *White Paper.*

Azevedo, H. L. C., Vellasco, M. M. B. R., & Passos, E. P. L. (2001). Data mining applied to CRM for a long distance telephone company. *Advances in Neural Networks and Applications*, 327-332.

Azzedin, F., & Maheswaran, M. (2002). Evolving and managing trust in Grid computing systems. In *Proceedings of the IEEE Canadian Conference on Electrical Computing Engineering (CCECE '02).*

Babbage, R., Moffat, I., O'Neill, A., & Sivaraj, S. (1997). Internet phone: Changing the telephony paradigm? *BT Technology Journal*, *15*(2), 145-157.

Baber, C. (2005). Evaluation of human-computer interaction. In J. R. Wilson, & N. Corlett, (Eds.), *Evaluation of Human Work*. 3rd edn. pp. 357-388. London, UK: Taylor & Francis.

Baccelli, E. (2007). *Address auto-configuration for MANET: Terminology and problem statement*. IETF draft, November 2007, draft-ietf-autoconf-statement-02.txt. Retrieved January 4, 2008, from http://tools.ietf.org/id/draft-ietf-autoconf-statement-02

Bacci, G., Luise, M., Poor, H. V., & Tulino, A. M. (2007). Energy Efficient Power Control in Impulse Radio UWB Wireless Networks. *IEEE Journal of Selected Topics in Signal Processing, 1*(3), 508-520.

Bakalis, S., Abeln, M., & Mante-Meijer, E. (1998) *Adoption and use of mobile telephony in Europe*. Cost 248 Report. Telia, Farsta, Sweden.

Baldwin, T., McVoy, D., & Steinfield, C. (1996). *Convergence: Integrating media, information and communication*. Thousand Oaks, CA: Sage.

Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The Smart Phone: A Ubiquitous Input Device. *Pervasive Computing, 5*(1), 70-77.

Ballon, P. (2007). Changing business models for Europe's mobile telecommunications industry: The impact of alternative wireless technologies. *Telematics and Informatics, 24*(3), 192-205.

Balzarova, M. A., Castka P., Bamber C. J., & Sharp J. M. (2006). How organisational culture impacts on the implementation of ISO 14001:1996 – a UK multiple-case view. *Journal of Manufacturing Technology Management, 17*(1), 89-103.

Banerjee, A., Madden, G., & Tan, J. (2007, January) *Regulatory trends: New enabling environment*. Paper presented at the ITU New Initiatives Programme workshop on "The Future of Voice", Geneva. Retrieved March 30, 2008, from http://www.itu.int/osg/spu/ni/voice/papers/FoV-Madden-Banerjee-Tan-Draft.pdf

Banker, R. D., Chang, H., Janakiraman, S. N., & Konstans, C., (2004). A balanced scorecard analysis of performance metrics. *European Journal of Operational Research, 154*, 423–436.

Banks, J., Olson, M., Porter, D., Rassenti, S., & Smith, V. (2003). Theory, experimental and the Federal Communications Commission spectrum auctions. *Journal of Economic Behaviour and Organization., 51*(3), 303-350.

Bansal, G., Hossain, J., & Bhargava, V. K. (2007). Adaptive power loading for OFDM-based cognitive radio systems. *IEEE International Conference on Communications, 1*, 5137-5142.

Barberis, G., & Pazzaglia, D. (1980). Analysis, design and buffer sizing of a packet voice receiver. *IEEE Transactions on Communications*, 217–227.

Barco, R., Díez, L., Wille, V., & Lázaro, P. (2008a). Automatic diagnosis of mobile communication networks under imprecise parameters. *Expert Systems with Applications.* In press.

Barco, R., Guerrero, R., Hylander, G., Nielsen, L., Partanen, M., & Patel, S. (2002b). Automated troubleshooting of mobile networks using Bayesian Networks. *Proceedings of the IASTED International Conference on Communication Systems and Networks (CSN'02).* Málaga, Spain (pp. 105-110)

Barco, R., Lázaro, P., Díez, L., & Wille, V. (2008b). Continuous versus discrete model in auto-diagnosis systems for wireless networks. *IEEE Transactions on Mobile Computing.* In press.

Barco, R., Nielsen, L., Guerrero, R., Hylander, G., & Patel, S. (2002a). Automated troubleshooting of a mobile communication network using Bayesian networks. *Proceedings of the IEEE International Workshop on Mobile and Wireless Communications Networks (MWCN'02).* Stockholm, Sweden (pp. 606-610).

Barco, R., Wille, V., & Díez, L. (2005). System for automated diagnosis in cellular networks based on performance indicators. *European Transactions on Telecommunications, 16*(5), (pp. 399-409).

Barco, R., Wille, V., Díez, L., & Lázaro P. (2006). Comparison of probabilistic models used for diagnosis in cellular networks. *Proceedings of the IEEE Vehicular Technology Conference VTC'06*. Melbourne, Australia.

Barney, J. B. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management, 17*(1), 99-120.

Barney, J. B. (1986). Strategic factor markets. Expectations, luck and business strategy, *Management Science, 32*, 1231-1241.

Barnum C., Bevan, N., Cockton. G., Nielsen, J., Spool, J., & Wixon, D. (2003). The "Magic Number 5": Is It Enough for Web Testing? In *Proceedings of CHI '03* (pp. 698-699). New York, NY: ACM Press.

Baron, D. P. (1993). *Business and its environment*. Englewood Cliffs, NJ: Prentice Hall.

Baron, D. P. (1995). Integrated strategy: Market and nonmarket components. *California Management Review*, *37*, 73-85.

Barreto, G. A., Mota, J. C. M., Souza, L. G. M., Frota, R. A., Aguayo, L., Yamamoto, J. S., & Oliveira, P. (2004). Competitive Neural Networks for Fault Detection and Diagnosis in 3G Cellular Systems. *Lecture Notes in Computer Science (LNCS), 3124*, 207-213.

Barreto, G. A., Mota, J. C. M., Souza, L. G. M., Frota, R. A., & Aguayo, L., Lehtimäki, P., Hätönen, K., & Simula, O. (2005). Condition monitoring of 3G cellular networks through competitive neural models. *IEEE Transactions on Neural Networks*, *16*(5), 1064-1075.

Barris, M. (2007). Cablevision Posts Wider Loss, Tempers Outlook for New Services. *The Wall Street Journal Online Edition*, November 8.

Bartlett, C. A., & Ghoshal, S. (1992). *Transnational Management: Text, Cases and Readings in Cross-Border Management* (3 ed.). Irwin: McGraw-Hill.

Baskerville, R., & Pries-Heje, J. (2001). A Multiple-Theory Analysis of a Diffusion of Information Technology Case. *Information Systems Journal,* (11), 3.

Bater, J., Tan, H.-P., Brown, K. N., & Doyle, L. (2007). Modelling interference temperature constraints for spectrum access in cognitive radio networks. *IEEE International Conference on Communications,1*, 6493-6498.

Bates, B., & Chambers, T. (2004). The cable industry. In A. Alison, J. Owers, R. Carveth, C. Hollifield, & A. Greco (Eds.), *Media economics: Theory and practice* (pp. 173-192). Mahwah, NJ: Lawrence Erlbaum & Associates.

Bates, T., Chandra, R., Katz, D., & Rekhter, Y. (2007). *Multiprotocol Extensions for BGP-4,* IETF, RFC 4760.

Bauer, J. M. (2003). Impact of license fees on the prices of mobile voice services. *Telecommunications Policy*, 27(5-6), 417-434.

Bauer, J. M., & Bohlin, E. (2008). From Static to Dynamic Regulation - Recent Developments in US Telecommunications Policy. *Intereconomics, 43*(1), 38-50.

Bauer, J., Kim, J., & Wildman, S. (2005). *Effects of national policy on the diffusion of broadband in OECD countries.* The UFL-LBS Workshop of the Future of Broadband: Wired and Wireless. Gainesville, FL, February 24-25, 2004.

BBC News. (2005). Legal fight hits music pirates. *British Broadcasting Corporation*. Retrieved November 14, 2005, from http://news.bbc.co.uk/1/hi/entertainment/music/4438324.stm

Beam, C., & Segev, A. (1997). Automated negotiations: A survey of the state of the art. *Wirtschafts - informatik, 39(3),* 263-268.

Beaubrun, R., Pierre, S., & Conan, J. (1999). An efficient Method for optimizing the assignment of cells to MSCs in PCS Networks. *Proceedings 11th International Conference On Wireless Communications (Wireless 99), 1*, 259-265.

Becher, R., Dillinger, M., Haardt, M., & Mohr, W. (2001). Broadband wireless access and future communication networks. *Proceedings of the IEEE*, *89*(1), 58-75.

Bedard, D. (2006). *The importance of IP address management*, Vaticor technical article, Retrieved April

6, 2007, from http://searchnetworking.techtarget.com/general/0,295582,sid7_gci1161808,00.html

Beer, B., d'Inverno, M., Luck, M., Jennings, N., Preist, K., & Schroeder, M. (1999). Negotiation in Multi-Agent Systems. *The Knowledge Engineering Review, 14(7),* 285-289.

Beijerse, U. R. P. (2000). Knowledge management in small and medium sized companies: Knowledge management for entrepreneurs. *Journal of Knowledge Management, 4*(2), 162-179.

Bell, J. (1995). The internationalization of small computer software firms: a further challenge to "stage" theories. *European Journal of Marketing, 29*(8), 60-75.

Benaroch M. (2002). Managing information technology investment risk: a real options perspective. *Journal of Management Information Systems*, *19*(2), 43-84.

Benaroch, M., & Kauffman, R. J. (1999). A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research, 10*(1), 70-86.

Benaroch, M., & Kauffman, R. J. (2000). Justifying electronic banking network expansion using real options analysis. *MIS Quarterly*, *24*(2), 197–225.

Benchaphon, L., Zhang, Y., & Zhang, Z. (2000). An Agent-based Negotiation Model supporting transactions in Electronic Commerce. *Proceedings of the 11th International Workshop on Database and Expert Systems Applications* (pp. 440 – 444). IEEE Computer Society, Washington, DC, USA.

Bendapudi, N., & Berry, L. L. (1997). Customers' motivations for maintaining relationships with service providers. *Journal of Retailing*, *73*(1), 15-37.

Benito, G. R. G., & Welch, L. S. (1994). Foreign Market Servicing: Beyond Choice of Entry Mode. *Journal of International Marketing, 2*(2), 7-27.

Benito, G. R. G., Larimo, J., Narula, R., & Pedersen, T. (2002). Multinational Enterprises from Small Economies. *International Studies of Management & Organization*, *32*(1), 57-78.

Bergh, D. D., & Holbein, G. F. (1997). Assessment and redirection of longitudinal analysis: Demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal, 18*(7), 557-571.

Bergman, E., & Norman, D. A. (2000). Making technology invisible: A conversation with Don Norman. In E. Bergman (Ed.), *Information Appliances and Beyond: Information Design for Consumer Products.* San Francisco, CA: Morgan Kaufmann.

Bernardos, C., Calderon, M., & Moustafa, H. (2007). *Survey of IP address autoconfiguration mechanisms for MANETs*, IETF draft, July 2007, draft-bernardos-manet-autoconf-survey-01.txt. Retrieved December 27, 2007, from http://ietfreport.isoc.org/idref/draft-bernardos-manet-autoconf-survey/.

Bernstein, A., & Goodman, S. (2005). Global Diffusion Of The Internet Vi: The Internet In Togo. *Communications of the Association for Information Systems, 15*(1), 371-392.

Berry, L. L. (2000). Cultivating service brand equity. *Journal of the Academy of Marketing Science, 28*(1), 128-37.

Bertolucci, J. (2007). Broadband EXPANDS. *PC World*, *25*(8), 76-88.

Bertsekas, D. (1998). *Network Optimization Continuous and Discrete Models.* Athena Scientific, Belmont, MA.

Bhagat, S., Bizjak, J., & Coles, J. L. (1998). The shareholder wealth implications of corporate lawsuits. *Financial Management*, *27*, 5-27.

Bharadwaj, S., Varadarajan. G. R., & Fahy, J. (1993). Sustainable competitive advantage in service industries: A conceptual model and research propositions. *Journal of Marketing*, *57*(4), 83-99.

Bhat, R. R. & Gupta, R. (2000). JAIN Protocol APIs. *IEEE Communications Magazine, 38*(1), 100-107.

Bhatt, G. D. (2000). Organising knowledge in the knowledge development cycle. *Journal of Knowledge Management*, *4*(1), 15-26.

Biermann, E., Cloete, E., & Venter, L. M. (2001). A comparison of intrusion detection systems. *Journal on Computers and Security, 20, I(8)*, 676-683.

Biersack, E., Carra, D., Lo Cigno, R., Rodriguez, P. and Felber, P. Overlay architectures for file distribution: Fundamental analysis for homogeneous & heterogeneous cases, Computer Networks 51 (2007) 901-917

Bigné, E., Ruiz, C., & Sanz, S. (2007). Key drivers of mobile commerce adoption an exploratory study of Spanish mobile users. *Journal of Theoretical and Applied Electronic Commerce Research*, *2*(2), 48-60.

Bilkey, W., & Tesar, G. (1977). The Export Behavior of Smaller-sized Wisconsin Manufacturing Firms. *Journal of International Business Studies, 8*(1), 93-98.

Bina, M., & Giaglis, G. M. (2005). *Exploring early usage patterns of mobile data services.* Paper presented at the International Conference on Mobile Business, Sydney, Australia.

Binmore, K. (2007). *Playing for Real. A Text on Game Theory*. New York, NY: Oxford University Press.

Binmore, K., & Vulkan, N. (1999). Applying game theory to automated negotiation. *Netnomics, 1(1)*, 1-9.

Birkan, G., Kennington, J., Olinick, E., Ortynski, A., & Spiride, G. (2006). Practical integrated design strategies for opaque and all-optical networks: optimization models and solution procedures. *Telecommuncation Systems , 31*(1), 61-83.

Bisdikian, C., Boamah, I., Castro, P., Misra, A., Rubas, J., Villoutreix, N., Yeh, D., Rasin, V., Huang, H., & Simonds, C. (2002). Intelligent Pervasive Middleware for Context-Based and Localized Telematics Services. *2nd International workshop on mobile commerce* (pp. 15-24). New York, NY: ACM Press.

Bizjak, J. M., & Coles, J. L. (1995). The Effect of Private Antitrust Litigation on the Stock-Market Valuation of the Firm. *The American Economic Review*, *85*, 436-461.

Black, U. (1995). *Network Management Standards: SNMP, CMIP, TMN, MIBs, and Object Libraries*. New York: McGraw-Hill Companies.

Blake, S. (Ed.). (1998). An Architecture for Differentiated Services, IETF, RFC 2475.

Bloemer, J., Brijis, T., & Vanhoof, K. (2002). Comparing complete and partial classification for identifying customers at risk. *International Journal of Research in Marketing*, *20*, 117–131.

Blum, R. (2006). *IP address management,* INS IT Industry Survey, October 2006. Retrieved February 15, 2007, from http://www.ins.com/assets/48615DD3-649A-4ED2-9C62-95196555966B.pdf

Blum, U., Growitsch, C., & Krap, N. (2007). Broadband Investment and the Threat of Regulation: Preventing Monopoly Exploitation or Infrastructure Construction? *Review of Network Economics*, *6*(3), 342-354.

Boddewyn, J. J., Halbrich, M. B., & Perry, A. C. (1986). Service Multinationals: Conceptualization, Measurement and Theory. *Journal of International Business Studies, 17*(3), 41-57.

Bodin L., Gordon L., & Loeb M. (2005). Evaluating information security investments using the analytic hierarchy process. *Communications of the ACM*, *48*(2), 78-83.

Bohannon, J. (2005). Grassroots Supercomputing. *Science*, *308*(5723), 810-813.

Bohn, R. E. (1994). Measuring and managing technological knowledge. *Sloan Management Review, 36*(1), 61-73.

Boles, C. (2007). AT&T Wins Partial Relief from FCC On Some Broadband Regulations. *The Wall Street Journal Online Edition*, Retrieved October 12, 2007.

Boleskei, H., Paulraj, A. J., Hari, K. V. S., & Naber, R. U. (2001). *Fixed Broadband Wireless Access: State of the Art, Challenges, and Future Directions*, IEEE Communications Magazine, (pp. 100-108). Jan, 2001.

Bollapragada, R., Morawski, T. B., Pinzon, L. E., Richman, S. H., & Sackett, R. (2007). Network planning of broadband wireless networks. *Interfaces, 37*(2), 143-162.

Bonald, T., & Proutière, A. (2003). Wireless Downlink Data Channels: User Performance and Cell Dimensioning. *Proceedings of ACM MobiCom'03*, September 14–19, 2003, San Diego, California, USA.

Bonardi, J. P. (1999). Market and nonmarket strategies of a former monopoly during deregulation: The British Telecom Case. *Business and Politics*, *1*, 203-232.

Bores, C., Saurina, C., & Torres, R. (2003). Technological convergence: A strategic perspective. *Technovation, 23*(1), 1-13.

Boston Globe. (2007). *Acton to vote on cell tower moratorium*.

Boucadair, M. (2005). *QoS-Enhanced Border Gateway Protocol*, draft-boucadair-qos-bgp-spec-01.txt.

Boucadair, M. (Ed.). (2006). *Parallel Internets Framework*, IST AGAVE project, deliverable D1.1, from http://www.ist-agave.org/results/D1.1-final-public.pdf.

Boulton, C. (2000). eBay joins mobile fray. *InternetNews.Com, May 22, 2000*. Retrieved on February 15, 2008, from http://www.internetnews.com/ec-news/article.php/376621

Boulton, R., Libert, B., & Samek, S.M.. (2000). Cracking the Value Code: How Successful Businesses Are Creating Wealth in the New Economy. *Harper Collins Publishers Inc*, New York.

Bourreau, M., & Dogan, P. (2001). Innovation and Regulation in the Telecommunications Industry. *Telecommunications Policy*, *25*(3), 167-184.

Bourreau, M., & Doğan, P. (2001). Regulation and innovation in the telecommunications industry. *Telecommunications Policy*, *25*(3), 167-184.

Bourreau, M., & Dogan, P. (2006). Build or Buy Strategies in the Local Loop. *American Economic Review, Papers and Proceedings , 96*(2), 72-76.

Boutremans, C., & Le Boudec, J. Y. (2003). Adaptive *playout* buffer and FEC adjustment for Internet Telephony. *In Proceedings of 22th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)* (pp. 652-662).

Bouwman, H., Haaker, T., & De Vos, H. (2007). Mobile service bundles: the example of navigation services. *Electronic Markets*, *17*(1), 20-28.

Bower, J. (1970). *Managing the Resource Allocation Process*. Homewood, IL.: Irwin.

Boyen, L. M. (2007, December). *RFC 5091 Identity-Based Cryptography Standard (IBCS)*.

Boylaud, O, & Nicoletti, G. (2001). Regulation, market structure and performance in telecommunications. *OECD Economic Studies*, *32*, 99-142.

Bracco, T. (2004). *Best practices in DNS/DHCP and IP address management*. NetworkWorld special report. Retrieved January 12, 2007 from http://www.genesiscom.ch/de/2/products/downloads/Lucent_IPAM-Bestpractises.pdf.

Brand, A. (1998). Knowledge management and innovation at 3M. *Journal of Knowledge Management*, *2*(1), 17-22.

Braunschweig, T., & Europe, N. C. (2003). *Jasmin: A Script-MIB Implementation*. From http://www.ibr.cs.tu-bs.de/projects/jasmin

Bräutigam, J., & Esche, C. (2003). Uncertainty as a key value driver of real options. *7th Annual Real Options Conference*.

Breese, J., & Heckerman, D. (1996). Topics in Decision-Theoretic Troubleshooting: Repair and Experiment. *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence* (pp. 124-132). Portland, Oregon.

Brewster, S. A. (2002). Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, *6*(3), 188-205.

Bria, A., Gessler, F., Queseth, O., Stridh, R., Unbehaun, M., Wu, J., & Zander, J. (2001). 4th-generation wireless infrastructures: scenarios and research challenges. *IEEE Personal Communications*, *8*(6), 25-31.

Brimson, J. A. (1991). *Activity accounting: An activity-based costing approach*. John Wiley & Sons, Inc.

Brimson, J. A., & Antos, J. (1994). *Activity-based management for service industries, governments entities and nonprofit organizations*. John Wiley & Sons, Inc.

Britton, B. B., & McGonegal, S. (2007) *The Digital Economy Fact Book, Ninth Edition 2007* (p. 61).Washington, D.C.: The Progress & Freedom Foundation.

Brix Networks. (2007). *Where does the 1-5 quality scale come from?* Retrieved November 24, 2007 from http://www.testmyvoip.com/faq.html#mos

Broadband Reports. (2007). *CODEC Choice in ViaTalk*. Retrieved November 27, 2007 from http://www.dslreports.com/forum/remark,18142109

Broadband Stakeholder Group (BSG). (2007). Pipe Dreams? *Prospects for next generation broadband deployment in the UK*. London: BSG.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland, (Eds.), *Usability Evaluation in Industry* (pp. 189-94). London, UK: Taylor & Francis.

Brooking, A. (1996). *Intellectual Capital: Core Assets for the Third Millennium Enterprise*. London: Thompson Business Press.

Brouthers, K. D., & Brouthers, L. E. (2003). Why Service and Manufacturing Entry Mode Choices Differ: The Influence of Transaction Cost Factors, Risk and Trust. *Journal of Management Studies*, *40*(5), 1179-1204.

Brown, A., & Patterson, D. (2000). Towards Availability Benchmarks: A Case Study of Software RAID Systems. *In Proceedings of 2000 USENIX Annual Technical Conference*.

Brumfield, C. (2007a). Time Warner Cable Falls to 50% Penetration Level. *IP Democracy*, November 8, 2007, Retrieved May 8, 2008 from http://www.ipdemocracy.com/archives/002755time_warner_cable_falls_to_50_penetration_level.php.

Brumfield, C. (2007b). Is Cable Telephony Growth Tapped Out? *IP Democracy*, November 15. Retrieved May 8, 2008 from http://www.ipdemocracy.com/archives/002765is_cable_telephony_growth_tapped_out.php.

Bruschi, D., Cavallaro, L., & Lanzi, A. (2007). An effective technique for preventing Mimicry and Impossible Paths Execution Attacks, *IEEE International Conference on Performance, Computing, and Communications*, (pp. 418-425). New Orleans, LA: Dipt. di Informatica e Comunicazione, Univ. degli Studi di Milano, Milan.

Bryan, P., & Adrian, P. (2005). **Challenges in securing vehicular networks**. In N. Spring (Ed.), *The Fourth Workshop on Hot Topics in Networks* (pp. 49-55). ACM SIGCOMM.

Buchanan, W. (1999). *Mastering Networks*. London: Macmillan Press Ltd.

Buckler, G. (2006). VoIP systems bring in new vulnerabilities. *Computer Dealer News*, *22*(9), 26.

Buckley, S. (2005). Governments grasp for IP communications. *Telecommunications (Americas Edition)*, *39*(11), 23-24.

Buckman, R. (1999). Collaborative knowledge. *Human Resource Planning*, *22*(1), 22-33.

Budenski, D. (2004). Keeping up with the evolution of telecommunications wiring. *EC and M: Electrical Construction and Maintenance*, *103*(8), 20-24.

Budiarjo, I., Nikookar, H., & Ligthart, L.P. (2006). Combined spectrum pooling and adaptive bit loading for cognitive radio OFDM-based system. *Symposium on Communications and Vehicular Technology, 1*, 73-76.

Buerkler, E. (2005). Dramatically different price developments for mobile services in India and Switzerland and their underlying drivers. *Info*, *7*(2), 72-75.

Bukowitz, W., & Williams, R.L. (2000). *The knowledge management fieldbook*. Revised edition, London: Prentice-Hall.

Burgelman, R. (1983a). A model of interaction of strategic behavior, corporate context, and the concept of strategy. *Academy of Management Review*, *3*(1), 61-69.

Burgelman, R. (1983b). A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly*(28), 223-244.

Burns, J., & Scapens, R. W. (2000). Conceptualizing management accounting change: an institutional framework. *Management Accounting Research*, *11*, 3-25.

Burnstein, D. E., & Aron, D. J. (2003). Broadband adoption in the United States: An empirical analysis. In A. L. Shampine (Ed.), *Down to the wire: Studies in the diffusion and regulation of telecommunications technologies (pp. 119-128).* New York: Nova Science Publishers.

Burs, M. J., & Woodruff, R. B. (1991). Value: An Integrative Perspective. *Proceedings of the Society for Consumer Psychology*, (pp. 59-64).

Burton, T. (2008). *Q4 consumer VoIP league table*. Retrieved March 9, 2008 from http://www.fiercevoip.com/story/q4-consumer-voip-league-table/2008-03-03

Burton-Jones, A., & Hubona, G. S. (2005). Individual Difference and Usage Behavior: Revisiting a Technology Acceptance Model Assumption. *The DATA BASE for Advances in Information Systems - Spring 2005,36*(2), 75.

Buzzacchi, L., Colombo, M., & Mariotti, S. (1995). Technological regimes and innovation in services: the case of the Italian banking industry. *Research Policy*, *24*, 151-168.

Byeon, S.-B (2006). *Customer Needs Analysis and Service Provisioning Method for the Home Network Service*. Korea Radio Station Management Agency.

Byeon, S.-B (2006). *Customer Needs Analysis and Service Provisioning Method for the Home Network Service*. Korea Radio Station Management Agency.

Bykowsky, M. (2003). A secondary market for the trading of spectrum: promoting market liquidity. *Telecommunications Policy*, 27(7), 533-541.

Cable & Wireless (2000). Competition in the Supply of IP-network. A Position Paper by Cable and Wireless.

Cabral, L. M. B., & Riordan, M. H. (1989). Incentives for cost reduction under price cap regulation. *Journal of Regulatory Economics, 1*(2), 93-102.

Cabric, D., Mishra, S. M., & Brodersen, R. W. (2004). Implementation issues in spectrum sensing for cognitive radios. *38th Asilomar Conference on Signals, Systems and Computers, 1*, 772-776.

Cabric, D., Tkachenko, A., & Brodersen, R. W. (2006). Spectrum sensing measurements of pilot, energy, and collaborative detection. *Military Communications Conference*, *1*, 1-7.

Cade M. (2003). *Global International Internet Capacity and Traffic Data*. ITU-T SG3 Joint Rapporteur Groups On Internet TFM & IIC Meeting of Apr. 2003.

Cadman, R., & Dineen, C. (2006). *European Telecom's Lost Investment: An analysis of the ECTA Scorecard*. Strategy and Policy Consultants Network Ltd (SPC Network).

Cagalj, M., Ganeriwal, S., Aad, I., & Hubaux, J.-P. (2005). On selfish behaviour in CSMA/CA networks. *24th Annual Joint Conference of the IEEE Computer and Communications Societies, 4*, 2513-2524.

Cahn, R. S. (1998). *Wide-Area Network Design: concepts and tools for optimization*. Morgan Kaufman Publishers, San Francisco, CA.

Cairncross, F. (1997). *The death of distance*. Boston: Harvard Business School Press.

Calhoun, P., Loughney, J., Guttman, E., Zorn, G., & Arkko J. (2003, September). *RFC 3588 Diameter Base Protocol*.

Camponovo, G., & Pigneur, Y. (2003). *Business Model Analysis Applied to Mobile Business.* Paper presented at the 5th International Conference on Enterprise Information Systems (ICEIS 2003), Angers, France, 23-26 April.

Cao, X., Annand, V., & Qiao, C. (2006). Framework for waveband switching in multigranular optical networks: part I - multigranular cross-connect architectures. *Journal of Optical Networking, 5*(12), 1043-1055.

Capron, L., & Chatain, O. (2008). Competitors´resource oriented strategies: Acting on competitors´resources through interventions in factor markets and political markets. *Academy of Management Review, 33*(1), 97-121.

Capuano, C. (2005). Access regulation in vertically integrated oligopolies. *Proceedings of 32th the European Association for Research in Industrial Economics (EARIE) Conference.*

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum.

Carlberg, K., & O'Hanlon, P. (2008, January). *RFC 5115 Telephony Routing over IP (TRIP).*

Carlberg, K., & O'Hanlon, P. (2008, January). *RFC 5115 Telephony Routing over IP (TRIP).*

Carnegie Mellon Software Engineering Institute (2007). *Common Management Information Protocol (CMIP).* Retrieved January 2008, from http://www.sei.cmu.edu/str/descriptions/cmip_body.html

Carr, H., & Snyder, C. (2002). *The Management of Telecommunications: Business Solutions to Business Problems.* New York: McGraw-Hill Companies.

Carroll, J. M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions.* Cambridge, MA: MIT Press.

Carson, D., Gilmore, A., Gronhaug, K., & Perry, C. (2001). *Qualitative Research in Marketing.* London: Sage.

Carvalho, R. P., & Alberti, A. M. (2007). Java technologies for NGN service creation: Discussion and architecture to improve SIP addresses discovery. *Internet and Multimedia Systems and Applications Conference*, (pp. 56 – 62).

Case, J. D., Fedor, M., Schoffstall, M. L., & Davin, J. (1989, April). RFC 1098 Simple Network Management Protocol (SNMP).

Case, J., Fedor, M., Schoffstall, M., & Davin, J. (1990). A Simple Network Management Protocol (SNMP). *Internet Engineering Task Force, Request for Comments: 1157 (STD: 15).*

Cassell, C., & Symon, G. (1994). *Qualitative Methods in Organisational Research*. London: Sage Publications.

Castells, M. (1996). *The rise of the network society*. Oxford: Blackwell.

Cava-Ferreruela, I., & Alabau-Munoz, A. (2006). Broadband policy assessment: A cross-national empirical analysis. *Telecommunications Policy*, article In Press.

Cave, M. (2004). Economic Aspects of the New Regulatory Regime for Electronic Communications Services. In P. Buiges & P. Rey (Eds), *The Economics of Antitrust and Regulation in Telecommunications* (pp. 27-41). E. Elgar.

Cave, M. (2006). Encouraging Infrastructure Competition through the Ladder of Investment. *Telecommunications Policy*, *30*(3-4), 223-237.

Cave, M. (2006). Encouraging network competition via the ladder of investment, *Telecommunications Policy*, *30*, 223-237.

Cave, M., & Vogelsang, I. (2003). How Access Pricing and Entry Interact. *Telecommunications Policy*, *27*, 717-727.

Cave, M., & Waverman, L. (1999). The transformation of international telecommunications. *Business Strategy Review*, *10*(1), 21-27.

Cave, M., Stumpf, U., & Valletti, T. (2006). *A Review of certain markets included in the Commission's Recommendation on Relevant Markets subject to ex ante Regulation*. Report for the European Commission.

Cavusgil, S. T. (1984). Differences Among Exporting Firms Based on Degree of Internationalization. *Journal of Business Research*, *12*(2), 195-208.

Cayrac, D., Dubois, D., Haziza, M., & Prade, H. (1994). Possibility theory in fault mode effects analyses - A satellite fault diagnosis application. *Proceedings of the IEEE International Conference on Fuzzy Systems* (pp. 1176-1181). Orlando, Florida.

Cellular News. (2006). *DOJ Approves AT&T-Bell South Merger.*

Cha, H., Park, J., & Kim, H. (2003). *Extended support for global connectivity for IPv6 mobile ad- hoc Networks.*

IETF draft, October 2003, draft-cha-manet-extended-support-globalv6-00.txt. Retrieved December 14, 2007, from http://tools.ietf.org/id/draft-cha-manet-extended-support-globalv6-00.

Chaffee, C. (2001). *Building the global fiber optics superhighway*. New York: Springer.

Chai, H. Z. (2006). Personalized Information Recommendation Service. *Journal of information, 9*, 118.

Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a framework for integrating agile development and user-centred design. In *Proceedings of Extreme Programming and Agile Processes in Software Engineering (XP06).*

Chamberland, S. (2004). An efficient heuristic for the expansion problem of cellular wireless networks. *Computers & Operations Research*, *31*(11), 1769-1791.

Chamberland, S., & Pierre, S. (2002). An efficient Algorithm for the Design Problem of Cellular Wireless Networks. *10th International Conference on Telecommunication Systems, Modeling and Analysis.*

Chan, L. L. M., Shaffer, M. A., & Snape, E. (2004). In search of sustained competitive advantage: the impact of organizational culture, competitive strategy, and human resource management practices on firm performance. *International Journal of Human Resource Management*, *15*(1), 17–35

Chandler, G. (1999). How to set up a forecasting system in telecommunications industry. *The Journal of Business Forecasting Methods & Systems, 18*(2), 3-6.

Chandra, A., Gong, W., & Shenoy, P. (2003). Dynamic Resource Allocation for Shared Data Centers Using Online Measurements. In *Proceedings of Eleventh International Conference on Quality of Service (IWQoS 2003).*

Chandra, R., & Scudder, J. (2002). *Capabilities Advertisement with BGP-4*, IETF, RFC 3392.

Chang, S. G., & Gavish, B. (1993). Telecommunications network topological design and capacity expansion: formulations and algorithms. *Telecommunications Systems*, *1*, 99-131.

Chan-Olmsted, S. (1998). Mergers, acquisitions, and convergence: The strategic alliances of broadcasting, cable television, and telephone services. *The Journal of Media Economics, 11*(3), 33-46.

Chany, A-M., Marras, W. S., & Burr, D. L. (2008). The effect of phone design on upper extremity discomfort and muscle fatigue. *Human Factors*, *49*(4), 602-618.

Charnsripinyo, C., & Tipper, D. (2003). Topological Design of Survivable Wireless Access Networks. *Design of Reliable Communication Networks (DRCN)* (pp. 371-378).

Chavez, P. M. (1996). Kasbah: An Agent Marketplace for Buying and Selling Goods. *Proceedings of the 1st International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM'96)* (pp. 75-90). Practical Application Company, London, UK.

Chavez., A., Dreilinger, D, Guttman, R., & Maes, P. (1997). A Real Life Experiment in creating an Agent Marketplace. *Proceedings of the 2nd International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, 1198,* 160-179. UK: Springer-Verlag London.

Chen C., Yu, C., Yang, S., & Chang, H. (2004). A customer-oriented service-enhancement system for the public sector. *Managing Service Quality, 14*(5), 414-425.

Chen, B., Rouskas, G., & Dutta, R. (2006). Clustering methods for hierarchical traffic grooming in large scale mesh WDM networks. *Proceedings of the 5th International Symposium on Communication Systems, Networks And Digital Signal Processing (CSNDSP)* .

Chen, B., Tong, L., & Varshney, P. K. (2006). Channel-aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine, 23*(4), 16-26.

Chen, D., & Waterman, D. (2007). Vertical Ownership, Program Network Carriage, and Tier Positioning in Cable Television: An Empirical Study. *Review of Industrial Organization,* 30(3), 227-251.

Chen, J. (2003). *Profiting from benchmarks: Reasons behind China's rush to develop standards.* From ChinaTechNews.com. April.

Chen, L. (2008). A model of consumer acceptance of mobile payment. *International Journal of Mobile Communications, 6*(1), 32-52.

Chen, W. T., Deng, Y-W., Huang, N. F., Wang, J. S., & Lin, H. C. (1999). On the design of a high-speed network environment for multimedia applications. *Proc. Natl. Sci. Counc. Repub. Of China Part A, 23*(1), 111-124.

Cheng, J., Yu, H.-C., & Sincoskie, W. (2005). Meeting the broadband access infrastructure demands: The promise of Gigabit Ethernet. *Technological Forecasting & Social Change, 72*(1), 1-10.

Chesbrough, H. W. (1999). The organizational impact of technological change: a comparative theory of national institutional factors. *Industrial and Corporate Change, 8*, 447-485.

Chesbrough, H. W. (2003). Environmental influences upon firm entry into new sub-markets: Evidence from the worldwide hard disk drive industry conditionally. *Research Policy, 32*(4), 659-678.

Chesher, M., & Kaura, R. (2000). The E-business (R) evolution, Hewlett-Packard Professional Books, from htttp://www.interhack.net/pubs/network-security/

Chevalier, A., & Kicka, M. (2006). Web designers and Web users: Influence of the ergonomics quality of the Web site on the information search. *International Journal of Human-Computer Studies, 64*(10), 1031-1048.

Chevreuil, A., Loubation, P., & Vandendorpe, L. (2000). Performance of general transmitter induced cyclostationarity precoders: analysis based on a MMSE-DF receiver. *IEEE Transactions on Signal Processing, 48*(11), 3072-3086.

Chiang, R. I. C., Rowe, G. B., & Sowerby, K. W. (2007). A quantitative analysis of spectral occupancy measurements for cognitive radio. *IEEE 65th Vehicular Technology Conference, 1*, 3016-3020).

Chien, H., & Wu. S. (2002). SODPM: a Sequence-oriented decision processmodel forunstructured group decision problems. *Behaviour& Information Technology, 21*(1), 59-69.

Chi-Ho, T., & Kwong, S. (2005). Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. *IEEE International Conference on Industrial Technology,* (pp. 51-56). Department of Computer Science, City University of Hong Kong, China.

Child, J., & Tse, D. (2001). China's transition and its implications for international business. *Journal of International Business Studies, 32*, 1-23.

China (2002). *Proposed Changes to ITU-T Recommendation D.50 on New Settlement Principles and Systems for the Internet.* ITU-T Study Group 3, Contribution 16, COM 3-16-E, April.

Chlamtac, I., Gumaste, & Szabó, Cs. (Eds.). (2005). *Broadband Services: Business Models and Technologies for Broadband Community Networks.* Wiley.

Cho, K., Fukuda, K., Esaki, H., Kato, A., The impact and implications of the growth in residential user-to-user traffic, ACM Sigcomm Conf., Pisa (2006) <www.acm.org/sigs/sigcomm/sigcomm2006>

Choi, C., & Shin, H. (1992). A comment on a model of vertical product differentiation. *The Journal of Industrial Economics, 40* (2), 229-231.

Chong, C. W., & Yeow, P. H. P. (2005). *An empirical study of perceived importance and actual implementation of knowledge management process in the Malaysian telecommunication industry.* Paper presented at the International Conference of Technology and Management: Challenges and Prospects. (pp. 182-192), Faculty of Business and Law, Multimedia University, Melaka.

Chong, C. W., Chong, S. C., & Wong, K. Y. (2007). Implementation of KM strategies in the Malaysian telecommunication industry: An empirical analysis. *VINE: The Journal of Information and Knowledge Management Systems, 37*(4), 452-470.

Chong, C. W., Chong, S. C., & Wong, K. Y. (in press). Is the Malaysian telecommunication industry ready for KM implementation. *Journal of Knowledge Management.*

Chong, C. W., Chong, S. C., & Yeow, P. H. P. (2006). KM implementation in Malaysian telecommunication industry: An empirical analysis. *Industrial Management & Data Systems*, *106(*8), 1112-1132.

Chong, S. C. (2006). KM implementation and its influence on performance: An empirical evidence from Malaysian Multimedia Super Corridor companies. *Journal of Information and Knowledge Management, 5*(1), 21-37.

Chong, S. C., & Choi, Y. S. (2005). Critical factors in the successful implementation of knowledge management. *Journal of Knowledge Management Practice, 5.* Retrieved February 13, 2006, from http://www.tlainc.com/articl90.htm.

Chou, C.-T., Sai, S. N., Hyoil, K, & Shin, K. G. (2007). What and how much to gain by spectrum agility*? IEEE Journal on Selected Areas in Communications, 25*(3), 576-588.

Choudhrie, J., Papazafeiropoulou, A., & Lee, H. (2003). A web of stakeholders and strategies: a case of broadband diffusion in South Korea. *Journal of Information Technology, 18*(4), 281-290.

Christensen, C. M. (1992). Exploring the Limits of the Technology S-Curve. *Production and Operations Management, 1*(4), 334-366.

Christensen, C. M. (1993). The Rigid Disk Drive Industry: A History of Commercial and Technological Turbulence. *Business History Review*, *67*(4), 531-588.

Christensen, C. M. (1997). *The Innovator's Dilemma*. Harvard Business School Press.

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston/Massachusetts: Harvard Business School Press.

Christensen, C. M., & Bower, J. (1996). Customer Power, Strategic Investment, and the Failure of Leading Firms. *Strategic Management Journal, 17,* 197-218.

Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution: Creating and sustaining successful growth*. Boston, Massachusetts: HBS Press.

Christensen, C. M., & Rosenbloom, R. S. (1995). Explaining the Attacker's Advantage: The Technological Paradigms, Organizational Dynamics, and the Value Network. *Research Policy, 24*, 233-257.

Christensen, C. M., Anthony, S. D., & Roth, E. A. (2004). *Seeing What's Next - Using the Theories of Innovation to Predict Industry Change*. Boston: HBS Press.

Christine, L., & Michel, B. (2007). Secure Anonymous Broadcasting in Vehicular Networks. In C. Ahlund & S. Cherkaoui (Eds.), *The First IEEE LCN Workshop On User MObility and VEhicular Networks (ON-MOVE)* (pp. 1300-1313), IEEE: Ireland.

Chu, K. F. (2003). An organizational culture and the empowerment for change in SMEs in the Hong Kong manufacturing industry. *Journal of Materials Processing Technology, 139*, 505–509

Chu, K. F. (2003). An organizational culture and the empowerment for change in SMEs in the Hong Kong manufacturing industry. *Journal of Materials Processing Technology 139,* 505–509.

Chuang, M. C., Chang, C. C., & Hsu, S. H. (2001). Perceptual factors underlying user preferences toward product form of mobile phones. *International Journal of Industrial Ergonomics*, *27*, 247-258.

Chun, W., Weiming, S., & Hamada, G., (2003). An adaptive negotiation framework for agent based dynamic manufacturing scheduling. *Proceedings of IEEE SMC 2003* (pp. 1211-1216). Washington DC.

Chunmei, Y., Mingchu, L., Jianbo, M., & Jizhou, S. (2004). Honeypot and scan detection in intrusion detection system, in *Proceedings of Canadian Conference on Electrical and Computer Engineering*, *2*, 1107–1110. School of Electronic Information Engineering, Tianjin University, China.

Cicora, W., Framer, J., & Large, D. (1999). *Modern cable television technology: Video, voice, and data communication*. San Francisco, CA: Morgan Kaufmann.

Ciena (2005). *The value of OTN for network convergence and IP/Ethernet Migration* (White paper) Ciena Corporation [Online:] www.xchangemag.com/bgpl/XC-ciena6.pdf

CISCO (2004). Increasing network availability, from *http://www.cico.com/wrap/public/779/largeent/learn/technologies/ina/IncreasingNetworksAvailability-Abstract.pdf.*

Cisco Systems, Global IP traffic forecast and methodology, White paper, <http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/net_implementation_white_paper0900aecd806a81aa.pdf>

CITEL (2001). *Response of Mexico to Base Questionnaire for Study of International Accords on Rates for Internet Service.*

Clark, A. (2001). Extensions to the E Model to incorporate the effects of time varying packet loss and recency. *ETSI T1A1.1/2001-037: Telchemy Incorporated.*

Clark, D. D., Lehr, W., Bauer, S. J., Faratin, P., Sami, R., & Wroclawski, J. (2006). Overlay Networks and Future of the Internet. *Journal of Communications and Strategies*, *3*(63), 1-21.

Clark, T., & Knowles, L. L. (2003). Global myopia: globalization theory in International Business. *Journal of International Management, 9*(4), 361-372.

Clark, T., Rajaratnam, D., & Smith, T. (1996). Toward a Theory of International Services: Marketing Intangibles in a World of Nations. *Journal of International Marketing, 4*(2), 9-28.

Clarke III, I. (2001). Emerging value propositions for m-commerce. *Journal of Business Strategies*, *18*(2), 133-148.

Clarke, L. (1994). *The Essence of Managing Change*, Englewood Cliffs, NJ: Prentice-Hall.

Clausen, T., & Baccelli, E. (2005). *Simple MANET address auto-configuration*. IETF draft, January 2005, draft-clausen-manet-address-autoconf-00.txt. Retrieved December 14, 2007 from http://www.tools.ietf.org/html/draft-clausen-manet-address-autoconf-00.

Claver, E., Gasco, J. L., Llopis, J., & Lopez, E. A., (2000). Analysis of a cultural change in a Spanish telecommunications firm, *Business Process Management, Journal, 6*(4).

Claver, E., Llopis, J., Garcia, D., & Molina, H. (1998). Organizational culture for innovation and new technological behaviour. *Journal of High Technology Management Research*, *9*, 55–68.

Clemens, N., & Rose, C. (2005). Intelligent power allocation strategies in an unlicensed spectrum. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 37-42.

CNET.com. (2002). *Old-fashioned corporate crackdowns in a brave new world*. Retrieved, from the World Wide Web: http://reviews.cnet.com/4520-3513_7-1027308-1.html.

Coase, R. (1959). The Federal Communication Commission. *The Journal of Law & Economics*, *2*, 1-40.

Coase, R. H. (1959). The Federal Communications Commission. *Journal of Law and Economics*, *2*, 1-40.

Cobb, I., Innes, J., & Mitchell, F. (1992). *Activity based costing: Problems in practice*. London: CIMA.

Cochennec, J. Y. (2002). Activities on next-generation networks under Global Information Infrastructure in ITU-T. *IEEE Communications Magazine, 40*(7), 98-101.

Cockton, G., & Woolrych, A. (2002). Sale must end: should discount methods be cleared off HCI's shelves? *Interactions, 9*(5), 13-18.

Cohen, B., Incentives build robustness in BitTorrent, <http://bitconjurer.org/BitTorrent/bittorrentecon .pdf> (2003)

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*, 128-152.

Cole, R., & Rosenbluth, J. (2001). Voice over IP performance monitoring. ACM Computer Communication Review, *31*(2), 9-24.

Collette, L. (2004). Cable telecommunications. In A. Grant and J. Meadows (Eds.), *Communication technology update*. NY: Focal Press.

Collins, J. C., & Porras, J. I. (1994). *Built to Last: Successful Habits of Visionary Companies*. London: HarperCollins.

Colombo, M. G., & Grilli, L. (2006). *Broadband access and broadband-based application: an empirical study on the determinants of adoption among Italian SMEs.* Working paper. Politecnico di Milano.

Colt Telecom (2002). *Annual Report*. Available at: www.colt.com.

Comisión del Mercado de las Telecomunicaciones (CMT). (2007). Consulta publica sobre redes de acceso de nueva generacion.

Contractor, F. J., & Kundu, S. K. (2000). Globalization of hotel services: an examination of ownership and alliance patterns in a maturing service sector. In Y. Aharoni & L. Nachum (Eds.), *Globalization of Services: Some implications for theory and practice*. London: Routledge.

Contractor, F. J., & Lorange, P. (2002). The growth of alliances in the knowledge-based economy. *International Business Review, 11*(4), 485-502.

Cook, R. A., & Lafferty, J. C. (1983). *Organizational Culture Inventory: OCI*. Plymouth, MI: Human Synergistics.

Cookson, M. D., & Smith, D. G. (2001). 3G Service Control. *BT Technology Journal*, *19*(1), 67-79.

Cooper, R. (1990). Cost classification in unit-based and activity-based manufacturing cost systems. *Journal of Cost Management*, *4*(3), Fall, 4-14.

Cooper, R. G. (1988). Predevelopment activities determine new product success. *Industrial Marketing Management*, *17*, 237–247.

Cooper, R., & Kaplan, R. S. (1991). Profit priorities from activity-based costing. *Harvard Business Review*, May-June, 130-135.

Cooper, R., & Kaplan, R. S. (1992). Activity-based systems: measuring the costs of resource usage. *Accounting Horizons*, *6*(3), 1-13.

Copeland, T. E., Weston, J. F., & Shastri, K. (2005). *Financial Theory and Corporate Policy*. Pearson Addison Wesley, Boston, MA.

Corbett, L. M., & Rastrick, K. N. (2000). Quality performance and organizational culture: A New Zealand study. *International Journal of Quality & Reliability Management, 17*(1), 14–26.

Cordeiro, C., Challapali, K., Birru, D., & Shankar, N. S. (2005). IEEE 802.22: the first worldwide wireless standard based on cognitive radios. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 328-337.

Corpus Christi Pioneers Metro-Wide Wi-Fi Mesh Network for AMR. (2007, July). *A Tropos Network Case Study*.

Corrocher, N. (2002). The diffusion of Internet telephony among consumers and firms: Current issues and future prospects. *Technological Forecasting and Social Change*, *70*, 525-544.

Corrocher, N., & Fontana, R. (2007). Expectations, Network Effects and Timing of Technology Adoption: Some Empirical Evidence from a Sample of SMEs in Italy. forthcoming in *Small Business of Economics.*

Coukos, E. D. (2001). *Knowledge management: Processes and strategies used in research universities*. Unpublished doctoral dissertation, Florida Atlantic University, Boca Raton, FL.

Cox, D., & Widom, J. (1998). The Sumatra project has been supported by the Centers for Telecommunications and Center for Integrated Systems at Stanford, by Pacific Bell, and currently by the National Science Foundation. Homepage: http://www-db.stanford.edu/sumatra

Cox, L. A., & Sanchez, J.R. (2000). Designing least-cost survivable wireless backhaul networks. *Journal of Heuristics, 6*(4), 525-540.

Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M., & Brown, B. (2006). Supporting ethnographic studies of ubiquitous computing in the wild. In *Proceedings of 2006 ACM Symposium on Designing Interactive Systems* (pp. 60-69). New York, NY: ACM Press.

Craig, C., & Graham, R. (2003). Rights management in the digital world. *Computer Law & Security Report, 19*(5), 356-362.

Cramton, P. (2002). Spectrum auctions. In Cave, M., Majumdar, S. & Vogelsang, I. (Eds.), *Handbook of Telecommunications Economics* (pp. 605-639). Amsterdam: Elsevier Science.

Cramton, P., Shoham, Y., & Steinberg, R. (2006). *Combinatorial auctions.* Cambridge: The MIT Press.

Crandall, B., Klein, G., & R.Hoffman, R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT press.

Crandall, R. (2007). Technical Difficulties. *The Wall Street Journal Online Edition*, November 8, 2007. Retrieved November 8, 2007.

Crandall, R. W., & Jackson, C. L. (2003). The $500 billion opportunity: The potential economic benefit of widespread diffusion of broadband internet access. In A. L. Shampine (Ed.), Down to the wire: Studies in the diffusion and regulation of telecommunications technologies *(pp. 155-193).* New York: Nova Science Publishers.

Crisp, J., & Elliot, B. (2005). *Introduction to fiber optics*, 3rd edition. London: Newnes.

Crosby, G. V., & Pissinou, N. (2007). Evolution of cooperation in multi-class wireless sensor networks. *32nd IEEE Conference on Local Computer Networks, 1*, 489-495.

Cuhls, K. (2003). From forecasting to foresight processes - new participative foresight activities in Germany. *Journal of Forecasting, 22*(2-3), 93-111.

Cukier K. N. (1997). *Peering and Fearing: ISP interconnection and Regulatory Issues,* Retrieved Nov. 10, 2003, from http://ksgwww.harvard.edu/iip/iicompol/Papers/Cukier.html.

D'Aveni, R. A. (1994). *Hypercompetition: Managing the Dynamics of Strategic Maneuvering.* New York, NY: The Free Press.

D'Aveni, R. A. (2002). Competitive Pressure Systems. *MIT Sloan Management Review, 44*(1), 39-49.

D'Halluin, Y., Forsyth, P. A., & Vetzal, K. R. (2007). Wireless Network Capacity Investment. *European Journal on Operational Research*, *176*(2007), 584-609.

Dacin, M. T., Oliver, C., & Roy, J. J. (2007). The legitimacy of strategic alliances: An institutional perspective. *Strategic Management Journal*, *28*, 169-187.

Dadashi, Y., Wilson, J. R., Sharples, S., & Clarke, T. (2008). Applications of Handheld Computers for the Rail Industry. In P. Bust (Ed.), *Contemporary Ergonomics: Proceedings of the Annual Ergonomics Conference, Nottingham, 2008*. London, UK: Taylor & Francis.

Daft, R. L. (2001). *Organisation Theory and Design*. Cincinatti, OH, South-Western College Publishing.

Daly, P. (1993). Navstar GPS and GLONASS - global satellite navigation systems. *Electronics & Communication Engineering Journal*, *5*(6), 349-357.

Damascio, M., & Ferreira, A. (2004). Interactive television usage and application: The Portuguese case study. *Computers & Graphics*, *28*, 139-148.

Damsgaard, J., & Lyytinen, K. (2001). The role of intermediating institutions in the diffusion of electronic data interchange (EDI): how industry associations intervened in Denmark, Finland, and Hong Kong. *The Information Society*, *17*(3), 195-210.

Danfeng, Y., & Fangchun, Y. (2007). A Universal Service Layer Management Model in NGN. *IEEE International Conference on Advanced Communication Technology*, 3, 1721-1725.

Daniel, E. M., & Grimshaw, D. J. (2002). An exploratory comparison of electronic commerce adoption in large and small enterprises. *Journal of Information Technology, 17*, 133-147.

Danneels, E. (2004). Disruptive technology Reconsidered: A Critique and Research Agenda. *Journal of Product Innovation Management, 21*(4), 246-258.

Datar, S., & Gupta, M. (1994). Aggregation, specification and measurement errors in product costing. *The Accounting Review, 69*(4), 567-591.

Datta, P., Masand, B., & Mani, D.R. (2001). Automated cellular modeling and prediction on a large scale, *Issues on the Application of Data Mining,* (pp. 485–502).

Davenport, T. (1993). *Process innovation*. Boston, MA. Harvard Business School Press.

Davenport, T., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston, Massachusetts: Harvard Business School Press.

Davenport, T., & Short, J. (Summer, 1990). The new industrial engineering: Information technology and business process redesign. *Sloan Management Review*, 11-27.

Davidson, J., Bhatia, M., Kalidinidi, S., Mukherjee, S., & Peters, J. (2006). *Voice over IP fundamentals* (2nd ed.): Cisco Press.

Davik, F., Yilmaz, M., Gjessing, S., & Uzun, N. (2004). IEEE 802.17 resilient packet ring tutorial. *IEEE Communications Magazine*, *42*(3), 112–118.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13,* 319-340.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science, 35, 982-1003.*

De Angelis, V., Izzo, L., Napolitano, A., & Tanda, M. (2005). Cyclostationarity-based parameter estimation of wide-band signals in mobile communications. *IEEE/SP 13th Workshop on Statistical Signal Processing, 1*,1038-1041.

de Chernatony, L., & Cottam S. (2006). Why are all financial services brands not great? *Journal of Product & Brand Management, 15/2* 88–97.

De Clercq, J., Ooms, D., Prevost, S. and Le Faucheur, F. Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers, IETF, RFC 4798 (2007)

De Figueirido, R. J. P., & Edwards, G. (2007). Does Private Money Buy Public Policy? Campaign Contributions and Regulatory Outcomes in Telecommunications. *Journal of Economics and Management Strategy*, *16*(3), 547-576.

De Long, D. W., & Fahey, L. (2000). Diagnosing cultural barrieres to knowledge management. *Academy of Management Executive, 14*(4), 113-127.

De, P. (2007). New methods for sensing bandlimited signals in cognitive radio. *IEEE 65th Vehicular Technology Conference*, *1*,1916-1920.

Deal, P., & Kennedy, A. (1982). *Corporate cultures.* Reading, MA: Addison-Wesley.

Dedicated Short Range Communications (2005), from http://www.itsa.org/itsa/files/pdf/DSRCICWhitePaper.pdf

Dedicated Short Range Communications (DSRC) Service (2004). *The FCC DSRC website*, http://wireless.fcc.gov/services/its/dsrc/

Deephouse, D. L., & Carter, S. M. (2005). An examination of differences between organizational legitimacy and organizational reputation. *Journal of Management Studies*, *42*(2), pp. 329-360.

Delaney, K., & Sharma, A. (2007). Google Has Even Bigger Plans for Mobile Phones. *The Wall Street Journal Online Edition*, November 16, 2007. Retrieved November 16, 2007.

DeLeon, P., & Sreenan, C. J. (1999). An adaptive predictor for media *playout* buffering. In ICASSP (1999), *Acoustics, Speech, and Signal IEEE Processing, 6*, 3097-3100.

Deleuze, G., & Guattari, F. (1996). *What Is Philosophy?* Columbia University Press ISBN: 0231079893.

Demarest, M. (1997). Understanding knowledge management. *Long Range Planning, 30*(3), 374-384.

Demichelis, C., & Chimento, P. (2002). *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM),* IETF, RFC 3393.

Deng, R. H., Lazar, A., & Wang, W. (1993). A Probabilistic Approach to Fault Diagnosis in Linear Lightware Networks. *IEEE Journal on Selected Areas in Communications*, *11*, 1438-1448.

Deng, T., & Subramaniam, S. (2003). Amplifier placement in transparent DWDM ring networks. In A. Somani, & Z. Zhang (Ed.), *Proceedings of SPIE*, *5285*, (pp. 246-257).

Denison, D. R. & Mishra, A. K. (1995). Toward a theory of organizational culture and effectiveness. *Organization Science*, *6*, 204-22.

Denison, D. R. (1984). Bringing corporate culture to the bottom line. *Organizational Behavior, 13*, 5-22.

Denison, D. R., Haaland, S., & Goelzer, P. (2004). Corporate culture and organizational effectiveness: Is Asia different from the rest of the world? *Organizational Dynamics, 33*, 98-109.

Denniston, F. (1998). FLAG – Fiber-Optic Link Around the Globe. *Sea Technology,* Feb., 78-83.

Denzin, N. K. (1989). *The Research Act: A Theoretical Introduction to Sociological Methods* (3rd ed.). Englewood Cliffs, N. J.: Prentice Hall.

Department of Justice (1999). *Justice Department Requires Bell Atlantic and GTE to Divest Wireless Businesses in order to Proceed with Merger.*

Department of Justice (2004). *Justice Department Requires Divestitures in Cingular Wireless's Acquisition of AT&T Wireless.*

DeSanctis, G., & Gallupe, B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(12), 1589-1609.

Deshpande, R., & Webster, F. E. Jr., (1989). Organizational culture and marketing: Defining the research agenda. *Journal of Marketing*, *53*, January, 3-15.

Deshpande, R., & Webster, F. E. Jr., (1989). Organizational culture and marketing: Definin the research agenda. *Journal of Marketing, 53*, January, 3-15

Deshpande, R., Farley, J. U., & Webster, F. E. (1993). Corporate culture, customer orientation and innovativeness in Japanese firms: a quadrad analysis. *Journal of Marketing, 57*, 23-7.

Dess, G. G., & Picken, J. C. (2000). Changing roles: leadership in the 21st century, *Organizational Dynamics*, 28(3), 18-34.

Deutsche Telekom (2001). Die Deutsche Telekom, France Telecom und Sprint sind von ihrem internationalen Joint Ventures Global One u¨berzeugt. Available at: www.telekom.de/dtag/presse/artikel/0,1018,x125,00.html (accessed 25 March).

Devarapalli, V. (2007, December). *RFC 5096 Mobile IPv6 Experimental Messages.*

Development Process: A new look. Washington, DC: National Academies Press.

Devroye, N., Mitran, P., & Tarokh, V. (2005). Cognitive multiple access networks. *Proceedings of IEEE International Symposium on Information Theory, 1*, 57-61.

Devroye, N., Mitran, P., & Tarokh, V. (2006). Achievable rates in cognitive radio channels. *IEEE Transactions on Information Theory, 52* (5), 1813-1827.

Dick, H., & Merrett, D. (2007). *The Internationalisation Strategies of Small Country Firms.* Cheltenham, UK: Edward Elgar.

Digital Spy. (2005). Report: Illegal file sharing on increase. *Digital Spy Limited*. Retrieved November 16, 2005, from http://www.digitalspy.co.uk/article/ds26272.html

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147 – 160.

Dimitrakopoulos, G., & Demestichas, P. (2005). Spectrum exchange in a reconfigurable radio context. *First IEEE*

International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1, 676-679.

Dimitriades, Z. S. (2005). Creating strategic capabilities: organizational learning and knowledge management in the new economy, *European Business Review, 17*(4), 314-324.

Dimitriades, Z. S., (2005). Creating strategic capabilities: organizational learning and knowledge management in the new economy. *European Business Review, 17*(4), 314-324

Distaso, W., Lupi, P., & Manenti, F. (2006). Platform competition and broadband uptake: Theory and empirical evidence from the European Union. *Information Economics and Policy, 18*(1), 87-106.

Dittrich, K., Jsapers, F., Valk, W., & Wynstra, F. (2006). Dealing with dualities. *Industrial Marketing Management, 35*, 792-796.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Dixit, A.K., & Pindyck, R.S. (1994). *Investment under uncertainty*, Princeton, NJ: Princeton University Press.

Dobson, G. (2002). Quality of service in service-oriented architectures, from http://digs.sourceforge.net/papers/qos.html.

Douskalis, B. (2000). *IP Telephony: The Integration of Robust VolP Services*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Doyle, C., & Smith, J. C. (1998). Market structure in mobile telecoms: qualified indirect access and the receiver pays principle. *Information Economics Policy*, 10(4), 471-488.

Dredden, G., & Bergdolt, J. (2007). Enterprise resource planning. *Air Force Journal of Logistics*, *2*(31), 48-52.

Drevers, T., va de Meent, R., & Pras, A. (2004). On the standardization of Web Services management operations. In *10th Open European Summer School and IFIP WG6.3 Workshop* (pp. 143-150).

Drezner, Z., & Hamacher, H. W. (2002). *Facility Location: Application and Theory*. Springer-Verlag, Heidelberg, Germany.

DSL-Forum, <www.dslforum.org/dslnews/pdfs/YE06_Release.pdf>

Du Gay, P. (1996). *Consumption and Identity at Work*. London, Sage.

Duffy, J. (2000). Knowledge management: To be or not to be? *Information Management Journal, 34*(1), 64-67.

Duh, H. B-L., Tan, G. C. B., & Chen, V. H. (2005). Usability evaluation for mobile device: a comparison of laboratory and field tests. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp 181-186). New York, NY: ACM Press.

Dunbar, J. (2006). *FCC Delays Vote on AT&T-BellSouth Deal*. Associated Press.

Dupont, J. K., Willers, J. L., Seal, M. R., & Hood, K. B. (1999). Precision pesticide applications using remote sensing. *Proc. 17-th Biennial Workshop on Color Photography and Videography in Resource Assessment, Reno, NV.*

Durand, P. (1983). The public service potential of videotext and teletext. *Telecommunications Policy*, *7*(6), 149–162.

Durham, D., Boyle, E. J., Cohen, R., Herzog, S., Rajan, R., & Sastry, A. (2000, January). *RFC 2748 The COPS (Common Open Policy Service) Protocol*.

Dutta, R., & Rouskas, N. (2002, November/December). Traffic grooming in WDM networks: past and future. *IEEE Network* (pp. 46-56).

Earston, A. (1980). Viewpoint. *Telecommunications Policy*, *4*(9), 220–225.

eBay (2006). *eBay Inc. Announces fourth quarter and full year 2005 financial results*. Retrieved March 31, 2007, from http://investor.ebay.com/news/Q405/EBAY0118-123321.pdf.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization*, *14*(2), 675-699.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization, 14*(6), 673-699.

Economides, N. (1999). The Telecommunications Act of 1996 and its impact. *Japan and the World Economy*, *11*, 455-483.

Edelman, L. B. (1992). Legal ambiguity and symbolic structures: organizational mediation of civil rights law. *American Journal of Sociology*, *97*(6), 1531–76.

Edwards, J. (2007). *VoIP QoS: What you need to know [Electronic Version]. VoIP News*. Retrieved December 2, 2007 from http://www.voip-news.com/feature/voip-qos-service-quality-012207/.

Edwards, P. N. et al. (2007, January). Understanding Infrastructure: Dynamics, Tensions, and Design. Report of a Workshop on *History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures*. NSF.

EIA/TIA, Technical Report IS-41, Revision B (1991). *Cellular Radio Telecommunications Intersystem Operations.*

Eiben, A. E., Raué, P. E., & Ruttkay, Z., (1995). *How to Apply Genetic Algorithms to Constrained Problems.* In L. Chambers (Ed.), *Practical Handbook of Genetic Algorithms, 1.* CRC Press, Inc.

Eisenhardt, K. M., & Brown, S. L. (1998). Competing on the edge: Strategy as structured chaos. *Long Range Planning, 31*(5), 786-9.

Eisenhart, K. M. (1989). Building theories from case research. *Academy of Management Review*, *14*(4), 532-550.

El Karoui, N., & Gobet, E. (2003). *Introduction au Calcul stochastique, première partie*, Imprimerie de l'Ecole Polytechnique.

El País (2006, November 20). *Las compañías telefónicas pierden siete de cada ocho pleitos judiciales contra el organismo regulador*, Madrid.

Elashmawi, F. (2000). Creating a winning corporate culture: Experience inside the Asian telecommunications industry. *European Business Review, 12*(3), 148–156.

ElBoghdady, D. (2001). Fiber-optic firms face issue of overbuilding. *Washington Post*, February 28, 1.

Electrical Engineering Times. (2007). FCC spells out rules for 700 MHz bands. *EE Times.*

Elixmann D., & Scanlan, M. (2002). *The Economics of IP Networks – Market, Technical and Public Policy Issues Relating to Internet Traffic Exchange*. wik-Consult.

Elliot, G., & Phillips, N. (2004). *Mobile Commerce and Wireless Computing Systems.* Upper Saddle River, NJ: Addison Wesley.

Elliott, I. (2008). *4Q07 consumer VoIP rankings*. Retrieved March 9, 2008 from http://ikeelliott.typepad.com/telecosm/2008/02/4q07-consumer-v.html

Enderlé, N., & Lagrange, X. (2003). User Satisfaction Models and Scheduling Algorithms for Packet-Switched Services in UMTS. *Proceedings of IEEE VTC spring*, April 2003.

Engebretson, J. (2007). Voice's place in the fiber future. *Telephony, 248*(9), 26-31.

Equipment%20Sept%202006%20AFTracker.pdf. (Verified January 23, 2008)

Erickson, T. (1995). Notes on design practice: stories and prototypes as catalysts for communication. In J. M. Carroll (Ed.), *Scenario-based design: Envisioning work technology in system development* (pp. 37-58). New York, NY: John Wiley & Sons.

Ericsson Jambala Parlay SCS, at: www.ericsson.com

Erlanger, L. (2005). Is VoIP ripe for attack? *InfoWorld*, *27*(42), 31.

Erramilli, M. K. (1990). Entry Mode Choice in Service Industries. *International Marketing Review, 7*(5/6), 50-62.

Erramilli, M. K., & Rao, C. P. (1993). Service firms' international entry-mode choice: A modified transaction-cost analysis approach. *Journal of Marketing*, *57*(3), 19-38.

Errol, E. J., & Winston B. E. (2005). A correlation of servant leadership, leader trust, and organizational trust. *Leadership & Organization Development Journal, 26*(1), 6-22.

Etkin, R., Parekh, A., & Tse, D. (2007). Spectrum sharing for unlicensed bands. *IEEE Journal on Selected Areas in Communications*, *25*(3), 517-528.

ETSI, Technical Report GSM, Digital Cellular Telecommunications System, Phase 2+ (1998). *Mobile Application Part (MAP) Specification (GSM 09.02 version 7.51 Release).*

EURESCOM project P1110 (2001). *Open Service Access: Advantages and opportunities in service provisioning on 3G Mobile Networks.* Document Proposal for Enhancements to the OSA/Parlay specifications.

EURESCOM, Traffic modelling for the new wireless services (TRAWIS) Eurescom study program P1443 (2004) 1-38 <http://www.eurescom.eu/Public/Projects/P1400-series/P1443/default.asp>

European city mayors population database. Retrieved May 15, 2005. (http://www.citymayors.com/features/euro_cities.html).

European Commission (2002). Guidelines 2002/C 165/03 on market analysis and the assessment of significant market power under the Community regulatory framework for electronic communications networks and services. *OJEC*, C165/31.

European Commission (1997). Notice on the definition of the relevant market for the purposes of Community competition law. *OJEC*, C 372.

European Commission. (2007). Proposal for a directive of the European Parliament and of the Council amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and services, and 2002/20/EC on the authorisation of electronic communications networks and services.

European Commission (2007). Recommendation 2007/879/EC of 17 December 2007 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services. *OJEU*, L344/65.

European Commission (2003). Recommendation 2003/311 of 11 February 2003 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services. *OJEC*, L 114/45.

European Regulators Group (ERG). (2004). ERG common position on the approach to appropriate remedies in the New Regulatory Framework. *ERG 30*(03).

European Regulators Group (ERG). (2007). ERG Consultation Document on Regulatory Principles of NGA, ERG (07) 16.

European Telecommunications Standards Institute (ETSI): Universal mobile telecommunication system (UMTS) – QoS concept and architecture, TS 123 107 / 3GPP TS 23.107 (2002) 1-37

European Union (1994). *Towards the Personal Communications Environment: Green Paper on a common approach to mobile and personal communications in the European Union*: COM (94) 145 final. Brussels, Commission of the European Communities.

European Union (2002). *On a common regulatory framework for electronic communications networks and service*. *Directive Framework* 2002/21/EC. Brussels, Commission of the European Communities.

European Union (2007). *European Electronic Communications Regulation and Markets 2006 (12th report).* (SEC[2007]403), Brussels, Commission of the European Communities.

European Union, Commission (1998). Declaring a Concentration to be Compatible with the Common Market. *Official Journal of the European Commission*.

European Union, Commission (2002). *Contribution from the European Commission regarding ITU-T Recommendation D.50 and Associated Work in Study Group 3*. ITU-T SG3 document COM 3-D 24-E, May.

Evans, P., Pucik, V., & Barsourx, J. L. (2002). *The Global Challenge: Frameworks for International Human Resource Management*. Boston, MA: McGraw-Hill Irwin.

Expert Choice (1995). *Expert Choice for windows based on Analytic Hierarchy Process*, Version 9, User Manual. Pittsburgh, PA.

Fadia, A. (2). (2006). *Network Security: A Hacker's Perspective*. Boston, MA: Thomson Course Technology.

Falch, M., & Tadayoni, R. (2004). Economic versus technical approaches to frequency management. *Telecommunications Policy*, *28*(2), 197-211.

Falch, M., & Tadayoni, R. (2007). Next generation broadband - Content and user perspectives. *Telematics and Informatics*, *24*(4), 243-245.

Falcione, R., Sussman, L., & Herden, R. (1987). Communication climate in organizations. In Handbook of Organizational Communication: An Interdisciplinary Perspective F. Jablin et al., (Eds.), Newbury Park, CA: Sage Publications, Inc.

Fan, Y., Saliba, A., Kendall, E. A., & Newmarch, J. (2005). Speech Interface: An Enhancer to the Acceptance of M-commerce Applications. *International Conference on Mobile Business*. Washington, DC: IEEE Computer Society.

Fang, T., Fridh, C., & Schultzberg, S., (2004). Why did the Telia–Telenor merger fail? *International Business Revie 13*, 573–594.

Faratin, P., Sierra, C., Jennings, N., Buckle, P. (1999). Designing Responsive and Deliberative Automated Negotiators. *Proceedings of the AAAI Workshop on Negotiation: Settling Conflicts and Identifying Opportunities* (pp 12–18). Carles Sierra's publications. Orlando, Florida, USA.

Fare, R., Grosskopf, S., & Knox Lovell, C. A. (1994). *Production Frontiers*. Cambridge University Press, Cambridge, Great Britain.

Farrell, J., & Saloner, G. (1985). Standardisation, compatibility and innovation. *Rand Journal of Economics*, *16*(1), 70–83.

Farrell, J., & Saloner, G. (1986). Installed base and compatibility: Innovation, product, preannouncement and predation. *American Economic Review*, *76*(5), 940–955.

Farrell, J., & Shapiro, C. (1990). Horizontal Mergers: An Equilibrium Analysis. *American Economic Review*, *80*(1), 107-123.

Farrell, J., & Weiser, P. (2003). Modularity, Vertical Integration, and Open Access Policies: Toward A Convergence of Antitrust And Regulation In The Internet Age. *Harvard Journal of Law & Technology, 17*(1), 85-134.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, *120*, 253-281.

Faulhaber, G. (2007). Network Neutrality: The Debate Evolves. *International Journal of Communications*, *1*(1), 680-700.

Faulhaber, G. R. (2002). Broadband deployment: Is policy in the way? In R. W. Crandall & J. H. Alleman (Eds.), *Broadband: Should we regulate high-speed Internet access?* (pp.223-244). Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Faulhaber, G. R., & Farber, D. J. (2003). Spectrum management: property rights, markets, and the Commons. In Cranor, L. F. & Wildman, S. S. (Eds.), *Rethinking rights and regulations: Institutional responses to new communication technologies* (pp. 193-226). Cambridge: The MIT Press.

Faycal, M. and Serhrouchni, A., CAP: A context-aware P2P system, Proc. OTM'07, Springer LNCS 4806 (2007) 950-959

Fayyad, M., & Irani, K.B. (1993). Multi-interval discretization of continuous valued attributes for classification learning. *Proceedings of the International Joint Conference on Artificial Intelligence* pp. 1022-1027). Chambery, France.

Fazio, M., Villari, M., & Puliafito, A. (2006). IP address auto-configuration in ad-hoc networks: Design, implementation and measurements, *Computer Networks*, *50*(7), 898-920.

FCC (2000). *The Digital Handshake: Connecting Internet Backbones*. OPP Working Paper No. 32.

FCC (1997). *Digital Tornado: The Internet and Telecommunication Policy*. OPP Working Paper No. 29.

FCC (1985). http://www.access.gpo.gov/nara/cfr/waisidx_01/47cfr15_01.html (Verified January 23, 2008)

FCC Federal Communications Commission (2004). ET Docket No. 03-237. *Establishment of an interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed, mobile and satellite frequency bands*.

FCC Federal Communications Commission (2005). ET Docket No. 03-208. *Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies*.

FCC News (1998). *Chairman William E. Kennard Announces Historic Agreement by Local and State Governments and Wireless Industries on Facilities Siting Issues*.

FCC News (2006). *FCC Approves Merger of AT&T Inc. and BellSouth Corporation*.

FCC News (1999). *FCC Approves SBC-Ameritech Merger Subject to Competition-Enhancing Conditions*

FCC News (2001). *FCC Announces Wireless Spectrum Cap to Sunset Effective January 1, 2003*.

Federal Communications Commission (2005). *Appropriate Framework for Broadband Access to the Internet over Wireline Facilities*. Policy Statement, CC Docket No. 02-33, FCC 05-151.

Federal Communications Commission (2007). *High-Speed Services for Internet Access: Status as of December 31, 2006*. Industry Analysis and Technology Division, Wireline Competition Bureau, October 2007.

Federal Communications Commission. (2000). *Fifth Annual CMRS Competition Report*.

Federal Communications Commission. (1999). *Fourth Annual CMRS Competition Report*.

Federal Communications Commission. (2004). *Ninth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services*.

Federal Communications Commission . (2007) *Second Report and Order, 700 Mhz Bands*.

Federal Communications Commission. (2002). *Seventh Annual CMRS Competition Report*.

Federal Communications Commission. (2005). *Tenth Annual Report and Analysis of Competitive Market Conditions with Respect to Commercial Mobile Services*.

Federal Communications Commission. (1998). *Third Annual CMRS Competition Report*.

Felegyhazi, M., Hubaux, J.-P., & Buttyan, L. (2005). Cooperative packet forwarding in multi-domain sensor networks. *3rd IEEE International Conference on Pervasive Computing and Communications Workshops*, *1*, 345-349.

Fellah, A. (2007). *WiMAX Market Trends and Deployments*. Paper presented at the WiMAX World Europe Vienna, Austria , May 29, 2007.

Fenner, B. (2003). *Smilint output for all MIBs published in RFCs and Internet Drafts*. From http://www.icir.org/fenner/mibs/htmllint

Fenwick, M., Edwards, R., & Buckley, P. J. (2003). Is cultural similarity misleading? The experience of Australian manufacturers in Britain. *International Business Review*, *12*(3), 297-309.

Fernández, Z., & Usero, B. (2007). The erosion of pioneer advantage in the European mobile telecommunications industry. *Service Business*, *1*(3), 195-210.

Ferreira, R., Gramma, A. and Jagannathan, S., Pletora: An efficient wide area storage system, Proc. HiCP'04, Springer LNCS 3296 (2004) 252-261

Fichman, R., Keil, M., & Tiwans, A. (2005). Beyond Valuation: "Option Thinking" in IT Project Management. *California Management Review*, *47*(2), 74-95.

Fidler, R. (1997). *Mediamorphosis: Understanding new media*. Thousand Oaks, CA: Pine Forge.

Field, L., Lowry, M., & Shu, S. (2005). Does disclosure deter or trigger litigation? *Journal of Accounting and Economics*, *39*, 487-507.

Fields, G. (2006). Innovation, time and territory: Space and the business organization of Dell Computer. *Economic Geography, 82*(2), 119-147.

Fife, E., & Pereira, F. (2005). *Global Acceptance of Technology (GAT) and Demand for Mobile Data Services By*. Retrieved July 4th 2008, from the World Wide Web: http://www.marshall.usc.edu/assets/004/5381.pdf

Figge, S. (2004). Situation-dependent services - a challenge for mobile operators. *Journal of Business Research, 57*(12), 1416-1422.

Financial Times (2008). *Sell phone*.

Fioreze, T., Granville, L. Z., Almeida, M. J., & Tarouco, L. R. (2005). Comparing Web Services with SNMP in a Management by Delegation Environment. In *IFIP/IEEE International Symposium on Integrated Network Management* (pp. 601-614).

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading MA: Addison-Wesley.

Fjeldstad, Ø., Becerra, M., & Narayanan, S. (2004). Strategic action in network industries: an empirical analysis of the European mobile phone industry. *Scandinavian Journal of Management*, *20*(1/2), 173-196.

Flamm, K. (2005, February). *The Role of Economics, Demographics, and State Policy in Broadband Availability*. Paper presented at the PURC/London Business School Conference on "The Future of Broadband: Wired and Wireless, 2005", Gainesville, FL.

Flanagin, A. J., & Metzger, M. J. (2001). Internet use in the contemporary media environment. *Human Communication Research*, *27*, 153-181.

Fletcher, R., & Fang, T. (2006). Assessing the impact of culture on relationship creation and network formation in emerging Asian markets. *European Journal of Marketing, 40*(3/4), 430-446

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177–192.

Florea, A., & Yanikomeroglu, H. (2005). On the optimal number of hops in infrastructure-based fixed relay networks. *IEEE Global Telecommunications Conference (GLOBECOM), 6*, 3242-3247.

Florian, D. (2005). Privacy issues in vehicular ad hoc networks. In George D. & David M. (Eds.), *The Fifth International Workshop On Privacy Enhancing Technologies* (pp. 197-209). Springer.

Floricel, S., & Miller, R. (2003). An exploratory comparison of the management of innovation in the New and Old economies. *R&D Management*, *33*(5), 501-525.

Fombrun, C. J. (1996). *Reputation: Realizing value from the corporate image*. Boston, MA: Harvard Business School Press.

FON. (2008). Retrieved January 30, 2008, from http://www.fon.com/

Fong, B., Ansari, N., Fong, A. C. M., Hong, G. Y., & Rapaji, P. B. (2004). On the scalability of fixed broadband wireless access network deployment. *IEEE Communications Magazine*, *42*(9), S12- S18.

Ford, G. S., & Spiwak, L. J. (2004). *The positive effects of unbundling on broadband deployment* (Policy paper No. 19). Washington, D.C.: Phoenix Center for Advanced Legal & Economic Public Policy Studies. Retrieved March 30, 2008, from http://www.phoenix-center.org/pcpp/PCPP19Final.pdf

Forman, H., Lippert, S. K., & Kothandaraman, K. (2007). Understanding users' performance evaluation of IT solutions. *Industrial Marketing Management*, *36*, 745-56.

Foster, W., Goodman, S., Osiakwan, E., & Bernstein, A. (2004). GLOBAL DIFFUSION OF THE INTERNET IV: THE INTERNET IN GHANA. *Communications of the Association for Information Systems, 13*(1), 654-681.

Foy, N. (1994). *Empowering People at Work*. Ashgate, Aldershot.

Frank, K., Zhendong M., & Elmar S. (2006). Security engineering for VANETs. In K. Borras (Ed.), *The Fourth Workshop on Embedded Security in Cars* (pp. 374-381). ACM: Germany.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research Methods in the Social Sciences* (5th ed.). New York: St. Martin's Press.

Fransman, M. (2002). Mapping the evolving telecoms industry: the uses and shortcomings of the layer model. *Telecommunications Policy, 26*(9/10).

Freeman, R. (2002). *Fiber optics systems for telecommunications.* New York: Wiley.

Frieden, S. (1999). Support services for distance education. *Journal of Educational Technology and Society, 2*(3), 77-88.

Friederiszick, H., Grajek, M., & Röller, L. H. (2008). *Analyzing the Relationship between Regulation and Investment in the Telecom Sector*. Paper presented at International Industrial Organization Conference, Marymount University, Arlington, VA.

Friedman, A. L., & Lyne, S. R. (1995). *Activity based techniques: The real life consequences*. London: CIMA.

Friedman, A. L., & Lyne, S. R. (1999). *Success and failure of activity-based techniques: A long term perspective*. London: CIMA;

Friedman, T. (2005). *The world is flat*. New York, NY: Farrar, Straus & Giroux.

Frischmann, B., & van Schewick, B. (in press). Network Neutrality and the Economics of an Information Superhighway: A Reply to Professor Yoo. *Jurimetrics, the Journal of Law, Science and Technology.*

Frommer, D. (2006). *Vonage dominates home VoIP market.* Retrieved March 9, 2008 from http://www.forbes.com/2006/03/14/voip-vonage-comcast-0314markets15.html

Frynas, E. G., Mellahi, K., & Pigman, G. A. (2006). First mover advantages in International Business and Firm-specific political resources. *Strategic Management Journal*, *27*, 321–345.

Fu, X., Zhou, W., Xu, J., & Song, J. (2007). Extended mobility management challenges over cellular networks combined with cognitive radio by using multi-hop network. *8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, *1*, 683-688.

Fujimoto, K., Ata, S., & Murata, M. (2001). *Playout* control for streaming applications by statistical delay analysis. *In Proceedings of IEEE International Conference on Communications (ICC 2001),8*, 2337-2342.

Fujimoto, K., Ata, S., & Murata, M. (2002). Adaptive *playout* buffer algorithm for enhancing perceived quality of streaming applications. *In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM 02)* (pp. 2463-2469), Taipei, Taiwan.

Fumagalli, A., Balestra, G., & Valcarenghi, L. (1998). Optimal amplifier placement in multiwavelength optical network based on simulated annealing. *SPIE International Symposium on Voice, Video, and Data Communicatios , 3531*, 268-279.

Funk, J. L. (2004). *Mobile Disruption - the technologies and applications driving the mobile internet.* Wiley.

Gabe, T., & Abel, J. (2002). Deployment of advanced telecommunications structure in rural America: Measuring the rural divide. *American Journal of Agricultural Economics, 84*, 1246-1252.

Gabrielsson, P., & Gabrielsson, M. (2004). Globalizing internationals: business portfolio and marketing strategies in the ICT field. *International Business Review, 13*(6), 661-684.

Galbraith, R., Turnbull, J., Dearnaley, R., Hart, J., & Nolde, K. (2005). Do we need quality of service in the customer environment? *BT Technology Journal, 23*(2), 146.

Galperin, H. (2005). Wireless networks and rural development: opportunities for latin america. *MIT Press Journal of Information Technologies and International Development*, *2*(3), 47-56.

Gancel, C., Rodgers, I., & Raynaud, M. (2002). *Successful Mergers, Acquisitions and Strategic Alliances: How to Bridge Corporate Cultures*. New York, NY: McGraw-Hill.

Gandetto, M., & Regazzoni, C. (2007). Spectrum sensing: A distributed approach for cognitive terminals. *IEEE Journal on Selected Areas in Communications, 25*(3), 546-557.

Ganesan, G., & Li, Y. (2007). Cooperative spectrum sensing in cognitive radio Part I: Two user networks. *IEEE Transactions on Wireless Communications, 6*(6), 2204-2213.

Ganesan, G., & Li, Y. (2007). Cooperative spectrum sensing in cognitive radio Part II: Multiuser networks. *IEEE Transactions on Wireless Communications, 6*(6), 2214-2222.

Gans, J. S. (2005). Concentration-based merger tests and vertical market structure. *Journal of Economic Literature*, *43*(2), 238-256.

Gao, M., Hyytinen, A., & Toivanen, O. (2006). Demand for and Pricing of Mobile Internet: Evidence from a Real-World Pricing Experiment. Helsinki Center of Economic Research Discussion Paper No. 122.

Gao, S., Qian, L., Vaman, D. R., & Qu, Q. (2007). Energy efficient adaptive modulation in wireless cognitive radio sensor networks. *IEEE International Conference on Communications*, *1*, 3980-3986.

Garavan, T. (1997). The learning organization: a review and evaluation, *The Learning Organization*, *4*(1), 18-29.

Garbin, D. A., & Gharakhanian, A. (2006). Voice quality in enterprise VOIP Systems. *Business Communications Review, 36*(2), 42.

Garcés-Erice, L., Biersack, E. W., Ross, K. W., Felber, P. A., & Urvoy-Keller, G. (2003). Hierarchical peer-to-peer systems. *Parallel Processing Letters*, *13*(4), 643-657.

Garcia, R. C., Sadiku, M. N. O., & Cannady, J. D. (2002). WAID: wavelet analysis intrusion detection. 45th. Midwest Symposium on *Circuits and Systems, 3,* 688–691. Science Atlanta, GA, USA.

García-Murillo, M. (2005). International broadband deployment: The impact of unbundling. *Communications & Strategies, 57*, 83-105.

García-Murillo, M., & Gabel, D. (2003, September). *International broadband deployment: The impact of unbundling.* Paper presented at the 31st Telecommunications Policy Research Conference, Washington, D.C.

Gardikis, G., Orfanos, S., Kormentzas, G., Pallis, E., & Kourtis, A. (2008). Dynamic IP configuration of terminals in broadcasting networks, *Computer Networks*, *52*(1). 292-302.

Gardner, W. A. (1986). The spectral correlation theory of cyclostationary time-series. *Signal Processing, 11*, 13-36.

Garg, V. K. (2002). *Wireless Network Evolution, 2G to 3G.* New Jersey: Prentice Hall.

Garretson, C. (2005). VoIP security threats: Fact or fiction? *Network World, 22*(29), 14.

Garretson, C. (2007). Hype vs. reality in VoIP security. *Network World, 24*(5), 20.

Garrett, J. J. (2003). *The Elements of User Experience*. Berkeley, CA: New Riders.

Gaspary, L., Barcellos, M., Detsch, A., Antunes, R. Flexible security in peer-to-peer applications: Enabling new opportunities beyond file sharing, Computer Networks 51 (2007) 4797-4815

Gatignon, H., & Xuereb, J. M. (1997). Strategic orientation of the firm and new product performance. *Journal of Marketing Research*, *34*(1), 77-90.

Gayle, P., & Weisman, D. (2007). Efficiency trade-offs in the design of competition policy for the telecommunications industry. *Review of Network Economics*, *6*(3), 321–341.

Gayle, P., & Weisman, D. (2007). Efficiency Trade-Offs in the Design of Competition Policy for the Telecommunications Industry. *Review of Network Economics*, *6*(3), 321-341.

Geir, J. (2002). *Wireless Lans, Implementing High Performance IEEE802.11 Networks.* Indianoplis, IN: SAMS Publishing.

Gelenbe, P. (2003). Will 'big brother' fears halt success of location-based mobile services. *Precision Marketing*, *15*(15), 16-17.

Gelli, G., Izzo, L., & Paura, L. (1996). Cyclostationarity-based signal detection and source location in non-Gaussian noise. *IEEE Transactions on Communications, 44*(3), 368-376.

George, S. (2002). Mobile Commerce: Vision and Challenges. *Proceedings of the Symposium on Applications and the Internet (SAINT.02.)* (pp. 43-45). IEEE Computer Society, Washington, DC, USA.

Georgiadis, C., Mavridis, I., & Manitsaris, A. (2005). Context-Based Humanized and Authorized Personalization in Mobile Commerce Applications. *International Journal of Computing & Information Sciences*, *3*(2), 1-9.

Geospatial Systems, Inc. (2008). http://www.geospatialsystems.com/imaging-products/ (Verified February, 6, 2008)

Gerding, E. H., van Bragt, D. D. B., & La Poutre, J. A. (2000). Scientific approaches and for negotiation. A game theoretic and artificial intelligence perspective. *CWI report on negotiation*, http://www.cwi.nl/projects/TA/.

Gerpott, T., Rams, W., & Schindler., A. (2001). Customer retention, loyalty, and satisfaction in the german mobile cellular telecommunications market. *Telecommunication Policy*, *25*(4), 249-269.

Gershman, A. (2002). Ubiquitous commerce – always on, always aware, always pro-active. *2002 Symposium on Applications and the Internet.* Washington, DC: IEEE Computer Society.

Gershon, R. (2000). *Telecommunications management: Industry structures and planning strategies.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gershon, R. (2002). The Transnational media corporation and the economics of global competition. In Yahya R. Kamalipour (Ed.), *Global Communication* (pp.51-73). Belmont CA: Wadsworth.

Gershon, R. A. (2004). Intelligent networking and the information economy. In K. Lewandowski (Ed.), *Annual Review of Communications*, *57*, 611-622. Chicago, IL: International Engineering Consortium.

Gershon, R., & Wirth, M. (1993). Home Box Office: The emergence of pay cable television. In R. Picard (Ed.), *The Cable Networks Handbook* (pp. 114-122). Riverside, CA: Carpelan Press.

Gerstner, L. (2002). *Who says elephants can't dance*? New York, NY: Harper Collins. Inc.

Ghosh, A., Wolter, D. R., Andrews, J. G., & Chen, R. (2005). Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential. *IEEE Communications Magazine, 43*(2), 129-136.

Gibson, J. (2000). *Multimedia Communications: Directions and Innovations.* Academic Press, 318 pages.

Gilbert, C. G. (2002). Beyond Resource Allocation: Towards a Process Model of Response to Disruptive Change. *Harvard Working Paper #03-018.*

Gilbert, C. G. (2005). Unbundling the Structure of Inertia: Resource versus Routine Rigidity. *Academy of Management Journal, 48*(5), 24.

Gilbert, C. G., & Bower, J. (2002). Disruptive Change - When Trying Harder Is Part of the Problem. *Harvard Business Review,* (5), 94-101.

Gilbert, L. A., & Kendall, J. (2003). *A Marketing Model for Mobile Data Services*. Paper presented at the Proceedings of HIICS-36, Honolulu, USA, 89-98.

Gilmore, T. N., Shea, G. P., & Useem, M. (1997). Side-effects of corporate cultural transformations. *Journal of Applied Behavioral Science*, *33*(2), 174-89.

Giordano, S., Salsano, S., Van den Berghe, S., Ventre, G., & Giannakopoulos, D. (2003). Advanced QoS provisioning in IP networks: The European Premium IP Projects. *IEEE Communications Magazine*, *41*(1), 30-36.

Glachant, J.-M. (2002). Why Regulate Deregulated Network Industries? *Journal of Network Industries, 3*, 297-311.

Gloet, M.,& Terziovski, M. (2004). Exploring the relationship between knowledge management practices and innovation performance. *Journal of Manufacturing Technology Management, 15*(5), 402–409.

Glover, F., & Laguna, M. (1997). *Tabu search*. Kluwer Academic Publisher.

Godfrey, J., & Kam, J. (2004). *What it takes to kick start mobile data: the lessons from Japan and South Korea*. ITU Asia 2004, Online News Service.

Goff, D. (2002). *Fiber optics reference guide*, 3rd edition. Woburn, MA: Focal Press.

Goggin, G., & Spurgeon, C. (2005). Mobile message services and communications policy. *Prometheus*, *23*(2), 181-193.

Goh, S. C. (2003). *Improving organizational learning capability: Lessons from two*.

Goldberg, D. E. (1989). Genetic algorithm in search, optimization and machine learning. New York: Addison-Wesley.

Goldsmith, M., Walt, C., & Doucet, K. (2000). New competencies for tomorrow's global leader, *CMA Management, 202473*(10), 20-4.

Golle, P., Greene, D. H., & Staddon, J. (2004). Detecting and correcting malicious data in VANETs. In H. Boral & P. Larson (Eds.), *The First ACM Workshop on Vehicular Ad Hoc Networks* (pp. 29-37). ACM: USA.

Gomez, C. (1999). *Engineering and Scientific Computing with Scilab*. ISBN: 978-0-8176-4009-5, 1999.

Goodman S, e. a. (1998). *The Global Diffusion of the Internet Project: An Initial Inductive Study,* . Retrieved July 2008, from the World Wide Web: http://mosaic.unomaha.edu/GDI1998/GDI1998.html

Gopal, A., Gopal, I., & Kutten, S. (1999). Fast broadcast in high-speed networks. *IEEE/ACM Transactions on Networking*, *7*(2), 262-275.

Gordon, G. G. (1991). Industry determinants of organizational culture. *The Academy of Management Review, 16*(2), 396-415.

Gordon, G. G., & DiTomaso, N. (1992). Predicting corporate performance from organizational culture. *Journal of Management Studies*, *29*, 783-798.

Goudos, S. K., Loutas, N., Peristeras, V., & Tarabanis, K. (2007). *Public Administration Domain Ontology for a Semantic Web Services EGovernment Framework*. Paper presented at the IEEE International Conference on Services Computing, (SCC 2007)

Gouveia, L., Patricio, P., & de Sousa, A. (2006). Hop-constrained node survivable network design: an application to MPLS over WDM. *Procedings of 8th INFORMS Teleommunication Conference.* Dallas, Texas.

Gouveia, L., Patricio, P., de Sousa, A., & Valadas, R. (2003). MPLS over WDM network design with packet level QoS constraints based on ILP models. *Proceedings of IEEE INFOCOM*.

Govindarajan, V., & Kopalle, P. K. (2006a). Disruptiveness of innovations: measurement and an assessment of reliability and validity. *Strategic Management Journal*, *27*, 11.

Govindarajan, V., & Kopalle, P. K. (2006b). The Usefulness of Measuring Disruptiveness of Innovations Ex Post in Making Ex Ante Predictions. *Journal of Product Innovation Management, 23*, 7.

Graham, S. (1999). Global grids of glass: On global cities, telecommunications, and planetary urban networks. *Urban Studies, 36*, 929-949.

Graham, S., & Marvin, S. (1996). *Telecommunications and the city: Electronic spaces, urban places*. London: Routledge.

Granstrand, O. (1994). Conclusions and Speculations. In E. Bohlin & O. Granstrand (Eds.), *The Race to European Eminence - Who are the coming Tele-Service Mulinationals?* (pp. 523-542). Amsterdam: Elsevier Science.

Grant, P. (2006). Cable Industry May need to Spend Heavily on Broadband Upgrades. *The Wall Street Journal*, August 17. p. A1.

Graps, A. (1995). An introduction to wavelets. *Computer Science and Engineering, 2, I(2),* 50-61. 22724 Majestic Oak Way, Cupertino, CA, USA.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Green, N., Harper, R. H. R., Murtagh, G., & Cooper, G. (2001). Configuring the mobile user: sociological and industry views. *Personal and Ubiquitous Computing, 5*, 146-156.

Greenstein, S., McMaster, S., & Spiller, P. (1995). The effect of incentive regulation of infrastructure modernization: Local exchange companies' deployment of digital technology. *Journal of Economics and Management Strategy, 4*(2), 187–236.

Greger K. R. (1999). A positive corporate culture is the soul of retention. *Hotel and Motel Management, 214*(17), 10.

Griss, M. L., & Pour, G. (2001). Accelerating development with agent components. *IEEE Computer mag., 34(5),* 37-43.

Grover, W. D., & Doucette, J. (2001). Topological Design of Survivable Mesh-Based Transport Networks. *Annals of Operations Research, 106*, 79-125.

Gruber, H., & Verboven, F. (2001). The Diffusion of Mobile Telecommunications Services in the European Union. *European Economic Review, 45*(577-588).

Gruber, J. G. (1981). Delay related issues in integrated voice and data networks. *IEEE Transactions on Communications, 29*, 786–800.

Grzybowski, L. (2005). Regulation of mobile telephony across the European Union: an empirical analysis. *Journal of Regulatory Economics, 28*(1), 47-67.

GSM Europe. (2008). 3GSM Licences: Overview of the Situation in the different EU Member States. *GSM World*, Retrieved February 15, 2008, from http://www.gsmworld.com/gsmeurope/faq/3g.shtml

Gu, Y., Towsley, D., Hollot, C., & Zhang, H. (2007). Congestion control for small buffer high-speed networks. *IEEE INFOCOM 2007* (pp. 1037-1045).

Gual, J., & Trillas, J. F. (2004). *Telecommunications policies: determinants and impact.* (August 2004). CEPR Discussion Paper No. 4578. Available at Social Science Research Network: http://ssrn.com/abstract=604468

Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.-J., & Nielsen, H. F. (2003). SOAP Version 1.2 Part 1: Messaging Framework. *W3C Recommendation*, from http://www.w3.org/TR/2003/REC-soap12-part1-20030624/

Guillory, B. (2006). Cable television. In A. Grant and J. Meadows (Eds.), *Communication technology update* (pp. 70-82). Boston, MA: Focal Press.

Gullati, R. (1998). Alliances and Networks. *Strategic Management Journal, 19*(4), 293-317.

Gumbley, H. (1998). Knowledge management. *Work Study, 47*(5), 175 -177.

Gunasekaran, A., Love, E., Rahimi, F., & Miele, R. (2001). A model for investment justification in information technology projects. *International Journal of Information Management, 21*, 349-364.

Gunasekaran, V., & Harmantzis, F. C. (2005). Affordable infrastructure for deploying WiMAX systems: mesh v.

non mesh. *IEEE 61st Vehicular Technology Conference*, (5), 2979- 2983.

Gunasekaran, V., & Harmantzis, F. C. (2007). Emerging wireless technologies for developing countries. *Technology in Society, 29*(1), 23-42.

Guthrie, G. (2006). Regulating Infrastructure: The Impact of Risk and Investment. *Journal of Economic Literature*, *XLIV,* 925-972.

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), 2902-2917.

Haddon, L. (2001). *Domestication and mobile telephony*. Presented at the conference "Machines that Become Us", Rutgers University, April 18-19, New Jersey.

Hagel, J., & Armstrong, A. G. (1997). *Net gain*. Boston, MA: Harvard Business School Press.

Häikiö, M. (2001). *The History of Nokia Oyj, Part III: Globalisation 1992–2000 (in Finnish)*. Helsinki: Edita Oyj.

Hakansson, H., & Ford, D. (2000). How should companies interact in business network? *Journal of Business Research*, *55*(2), 133-139.

Hale, A. (2000). Culture's confusions. *Safety Science, 34,* 1–14.

Haley, B., Devarappali, V., Deng, H., & Kempf, J. (2008, January). *RFC 5142 Mobility Header Home Agent Switch Message*.

Hall, B. H., & Khan, B. (2003). *Adoption of New Technology*. Working paper. Institute of Business and Economic Research, University of California, Berkley.

Hallikainen, P., Kivijarvi, H., & Nurmimaki, K., (2002). Evaluating strategic IT investments: An assesment of investment alternatives for a web content management system. *Proceedings of the 35th Hawaii International Conference on Systems Sciences.*

Halonen, T., Romero, J., & Melero, J. (2003). *GSM, GPRS & EDGE Performance Evolution Towards 3G/UMTS*. Chichester, England: Wiley.

Hamblin, R. L., Miller, J. L. L., & Saxton, D. E. (1979). Modeling use diffusion. *Social Forces, 57*(3), 799-811.

Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*. New York, NY: Harper Business.

Hammer, M., & Stanton, S. (1995). *The Reengineering Revolution: A Handbook***.** New York, NY: Harper Business.

Han, S., Choi, S. H., Kim, B., & Chung, H. (2006). A quantitative analysis of the effects of a handset subsidy on consumer welfare. *Electronics and Telecommunications Research Institute (ETRI) Journal, 28*(5), 621-630.

Handley, M., Jacobson, V., Perkins, C. (2006). *SDP: Session Description Protocol*. Request for Comments (Standards Track), 4566, Internet Engineering Task Force.

Handley, M., Schulzrinne, H., Schooler, E., & Rosenberg, J. (1999, March). *RFC 2543 SIP: Session Initiation Protocol*.

Hankinson, H., & Lloyd, S. P. (1993). Change Management. In Brinker, B. J. (Ed.), *Handbook of cost management*. New York: Warren Gorham Lamont.

Hankinson, P., & Hankinson, G. (1999). Managing successful brands: An empirical study which compares the corporate cultures of companies managing the world's top 100 brands with those managing outsider brands. *Journal of Marketing Management, 15*, 135-55.

Hannan, T. H., & McDowell, J. M. (1984). The determinants of technology adoption: the case of the banking firm. *RAND Journal of Economics, 15*, 328-335.

Harmatos, J. (2002). Planning of UMTS Core Network. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2,* 740-744.

Harmatos, J., Juttner, A., & Szentesi, A. (1999). Cost-based UMTS Transport Network Topology Optimisation. *International Conference on Computer Communication* (I*CCC'99).*

Harmatos, J., Szentesi, A., & Godor, I. (2000). Planning of Tree-Topology UMTS Terrestrial Access Networks.

11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1, 353–357.

Harrington, D., Presuhn, R., & Wijnen, B. (2002). An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. *Internet Engineering Task Force, Rquest for Comments: 3411 (STD: 62).*

Harris, L. C., & Ogbonna, E. (2001). Leadership style and market orientation: An empirical study. *European Journal of Management*, 32(5/6), 744-764.

Harris, N. (2000). The Basic Question Which wireless standard is best? A lot hangs in the balance. *The Wall Street Journal.*

Harrison, J. S., Hitt, M. A., Hoskisson, R. E., & Ireland, R. D. (1991). Resource Complementarity in Business Combination: Extending the Logic to Organizational Alliances, *Journal of Management*, *27*(6), 173-190.

Harte, L. (2007). *Introduction to cable television*: *Analog and digital television and modems.* Fuquay Varina, NC: Althos Publishing.

Hartleb, F. and Haßlinger, G. Comparison of link dimensioning methods for TCP/IP networks, Proc. IEEE GLOBECOM, (2001) 2240-47

Harvard Law Review (Eds) (2001). Antitrust and the Information Age:Section 2 Monopolization Analyses in the New Economy. *Harvard Law Review, 114*, 5.

Haßlinger, G. Quality-of-service analysis for statistical multiplexing with Gaussian and autoregressive input modeling, Telecommunication Systems 16 (2001) 315-334

Hasslinger, G., Guillemin, F., Ferreira J. and Halldorsson, U. The impact of peer-to-peer networking on network operators and Internet service providers, Eurescom study report P1553 (2005) <www.eurescom.eu/public/projects/P1500-series/p1553/>

Haßlinger, G., Mende, J., Geib, R., Beckhaus, T. and Hartleb F.: Measurement and characteristics of aggregated traffic in broadband access networks, Proc. 20. Internat. Teletraffic Congress, Ottawa, Canada, Springer, LNCS 4516 (2007) 998-1010

Hasslinger, G. ISP Platforms under a heavy peer-to-peer workload, Proc. Peer-to-Peer Systems and Applications, Eds.: R. Steinmetz and K. Wehrle, Springer LNCS 3485 (2005) 369-382

Haßlinger, G., Schnitter, S. and Franzke, M. The Efficiency of Traffic Engineering with Regard to Failure Resilience, Telecommunication Systems Vol. 29/2, Springer (2005) 109-130

Hausman, J. (2002). Internet-related services: The results of asymmetric regulation. In R. W. Crandall & J. H. Alleman (Eds.), *Broadband: Should we regulate high-speed Internet access?* (pp.129-156). Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Hausman, J. (1997). Valuing the effect of regulation on new services in telecommunications. *Brookings Papers on Economic Activity, Microeconomics*, *28*(1), 1-54.

Hausman, J., & Sidak, G. (2005). Did Mandatory Unbundling Achieve its Purpose? Empirical Evidence From Five Countries. *Journal of Competition Law and Economics*, *1*, 173–245.

Hausman, J., & Sidak, G. (2005). Did mandatory unbundling achieve its purpose? Empirical evidence from five countries. *Journal of Competition Law and Economics, 1*(1), 173–245.

Havila, V., Johansson, J., & Thilenius, P. (2004). International business – relationship triads. *International Marketing Review*, *21*(2), 172-186.

Hawawini, G., & Viallet, C. (2007). *Finance for executives – managing for value creation*. Mason, Ohio: Thomson South-Western.

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications, 23*, 201-220.

Haykin, S., Huber, K., & Chen, Z. (2004). Bayesian sequential state estimation for MIMO wireless communications. *Proceedings of IEEE (Special Issue on Sequential State Estimation)*, *92*(3), 439-454.

Hazlett, T. W. (1998). Assigning Property Rights to Radio Spectrum Users: Why Did FCC License Auctions Take 67 Years? *Journal of Law and Economics*, *41*, 45-67.

Hazlett, T. W. (2003). Liberalizing US spectrum allocation, *Telecommunications Policy*, *27*(7), 485-499.

Hazlett, T. W., & Bazelon, C. (2005). *Regulated Unbundling of Telecommunications Networks: A Stepping Stone to Facilities-Based Competition?* Paper presented to the Telecommunications Policy Research Conference.

Hazlett, T., & Bazelon, C. (2005). Regulated unbundling of telecommunications networks: A stepping stone to facilities-based competition? Paper presented at the 33rd Telecommunications Policy Research Conference, Arlington, VA.

HBO (1984). *Pay TV guide: Editor's pay TV handbook*. New York: NY

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal*, *3*(3), 118-126.

Hecht, J. (1999). *City of light: The story of fiber optics*. Oxford and New York: Oxford University Press.

Heckerman, D., Breese, J., & Rommelse, K. (1995). Decision-theoretic troubleshooting. *Communication of the ACM*, *38*(3), 49-57.

Heckerman, D., Horvitz, E., & Nathwani, B. (1992). Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine*, 31, 90-105.

Hedible, C., & Pierre, S. (2003). A Genetic Algorithm for Assigning Cells to Switches in Personal Communication Networks. *IEEE Canadian Review*, *44*, 21-24.

Heijden van der, H., Ogertschnig, M., & van der Gaast, L. (2005). *Effects of Context Relevance and Perceived Risk on User Acceptance of Mobile Information Services*. ECIS 2005.

Heinonen, K., & Pura, M. (2006). Classifying Mobile Services. *Presented at Helsinki Mobility Roundtable*, 1-2 June 2006, Helsinki.

Helgstrand, K. K., & Stuhlmacher, A. F. (1999). National culture: An influence on leader evaluations? *The International Journal of Organizational Analysis*, *7*, 153-68.

Hellenthal, J. W., Panken, F. J. M., & Wegdam, M. (2001). Validation of the Parlay API through prototyping. *IEEE Intelligent Network Workshop.*

Heng, T. M., & Low, L. (1990). Towards greater competition in Singapore's telecommunications. *Telecommunications Policy*, *14*(4), 303-314.

Hennart, J. F. (1988). A transaction cost theory of equity joint ventures. *Strategic Management Journal*, *9*(4), 361-74.

Hicks, J. E., Mackenzie, A. B., Neel, J. A., & Reed, J. H. (2004). A game theory perspective on interference avoidance. *IEEE Global Telecommunications Conference*, *1*, 257-261.

Hilas, C. S., Goudos, S. K., & Sahalos, J. N. (2006). Seasonal decomposition and forecasting of telecommunication data: A comparative case study. *Technological Forecasting and Social Change*, *73*(5), 495-509.

Hill, T. (1985). *Manufacturing Strategy*, Basingstoke Macmillan.

Hillman, A. J., & Hitt, M. A. (1999). Corporate Political Strategy Formulation: A Model of Approach, Participation, and Strategy Decisions. *The Academy of Management Review*, *24*(4), 825-842.

Hillman, A. J., & Wan, W. P. (2005). The determinants' of MNE subsidiaries' political strategies: Evidence of institutional duality. *Journal of International Business Studies*, *36*, 322-340.

Hodendorn, C. (2005). Regulating vertical integration in broadband: Open access versus common carriage. *Review of Network Economics*, *4*(1), 19-32.

Hoekman, B. (1995). *Trade laws and institutions: Good practices and the World Trade Organization*. Supplementary paper presented at The Uruguay Round and Developing Economies, *a World Bank Conference*, 26-27 January 1995.

Hoff, R. (2004, October 11). Building an idea factory. *Business Week* (p. 194).

Hoffman, N., & Klepper, R. (2000). Assimilating new technologies. The role of organizational culture. *Information Systems Management, 17*(3), 36-42.

Hoffman, W. (2007, February 26). Dell's logistics restart. *Traffic World* (p. 1).

Höflich, J. R., & Rössler, P. (2001). Mobile schriftliche Kommunikation oder: E -Mail für das Handy. *Medien & Kommunikationswissenschaft*, *49*, 437-461.

Hofmann, P. (2006). *Multi-hop Radio Access Network (MRAN) Protocol Specification*. IETF draft, March 2006, draft-hofmann-autoconf-mran-00.txt. Retrieved December 5, 2007, from http://www.tools.ietf.org/html/draft-hofmann-autoconf-mran-00

Hofstede, G. (1991). *Cultures and Organisations: Software of the Mind*. New York, NY,McGraw-Hill.

Hofstede, G. (1993). Cultural Constraints in Management Theories. *Academy of Management Review,, 7*(1), 81-94.

Hofstede, G., (1994). The business of international business is culture, *International Business Review*, *3*(1), 1-14.

Hoglund, A. J., Hatonen, K., & Sorvari A. S. (2000). A computer host-based user anomaly detection system using the self-organizing map. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 5*, 411-416. Como, Italy.

Holger, F., Marc, T. M., Matthias, T., Roland, K., Hannes, H., & Wolfgang, E. (2005). Studying vehicle movements on highways and their impact on ad-hoc connectivity. In M. Satyanarayanan (Ed.), *The Eleventh Annual International Conference on Mobile Computing and Networking* (pp. 26-27). ACM: Germany.

Holland, J. (1985). *Making Vocational Choices*, 2nd ed., Englewood Cliffs, NJ Prentice-Hall.

Holleis, P., Kern, D., & Schmidt, A. (2007). Integrating user performance time models in the design of tangible UIs. *Proceedings of CHI'07 extended abstracts.* New York, NY: ACM Press.

Holsapple, C. W., & Joshi, K. D. (2001). Organizational knowledge resources. *Decision Support Systems, 31*(1) 39-54.

Hopper, T. (1994). Activity based costing: A critical commentary. *Fukuoka University Review of Commercial Sciences*, *39*(1-2), 479-511.

Hopper, T., & Major, M. (2007). Extending institutional analysis through theoretical triangulation: Regulation and activity-based costing in Portuguese telecommunications. *The European Accounting Review*, *16*(1), 59-97.

Hopper, T., Kirkham, L., Scapens, R., & Turley, S. (1992). Does financial accounting dominate management accounting – A research note. *Management Accounting Research*, *3*, 307-311.

Hosia, A. (2003). *Comparison between Radius and Diameter (T-110.551 Seminar on Internetworking).* Helsinki, University of Helsinki, Department of Computer Science.

House, R., Javidan M., Hanges, P., & Dorfman, P., (2002). Understanding cultures and implicit leadership theories across the globe. *Journal of World Business*, *37*, 3-10.

Housley, R. (2007, November*). RFC 5083 Cryptographic Message Syntax (CMS) Authenticated-Enveloped-Data Content Type.*

Housley, R. (2007, November*). RFC 5083 Cryptographic Message Syntax (CMS) Authenticated-Enveloped-Data Content Type.*

Howard, J., Intven, H., & Zohar, M. (1998), *Internet Telephony – The Regulatory Issues*. McCarthy Tetrault Publications.

Howell, B. (2002). *Broadband uptake and infrastructure regulation: Evidence from the OECD countries* (Working Paper BH01/01). Wellington: New Zealand Institute for the Study of Competition and Regulation.

Howell, M., Love, S., & Turner, M. (2006). Visualisation improves the usability of voice-operated mobile phone services. *International Journal of Human-Computer Studies, 64*(8), 754-769.

Hsieh, C., Jones, C., & Lin, B. (2008). The new business potential with mobile commerce. *International Journal of Mobile Communications, 6*(4), 436-455.

Hu, W.-M., & Prieger, J. (2008). The timing of broadband provision: The role of competition and demographics. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission* (pp. 241-259). Hershey, PA: IGI Global.

Huber, G. (1990). A theory of the effects of advanced information technologies on organizational design, intelligence and decisionmaking. *Academy of Management Review*, *15*, 195-204.

Huber, G. B. (2001). Transfer of knowledge in knowledge management systems: Unexplored issues and suggested studies. *European Journal of Information Systems, 10*(2), 72-79.

Hunt, G., & Arden, P. (2005). QoS requirements for a voice-over-IP PSTN. *BT Technology Journal, 23*(2), 37.

Hunt, S. D. (1999). *A general theory of competition: resources, competences, productivity, economic growth*. London: Sage Publications.

Hurley, S. (2002). Planning Effective Cellular Mobile Radio Networks. *IEEE Transactions on Vehicular Technology, 51*(2), 243-253.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38*, 635-72.

Hüsig, S., Dowling, M., & Hipp, C. (2007, 1-3 June). *Analysing the Disruptive Potential of WLAN: A Three-Country Comparison of the Hotspot Market.* Paper presented at the 5th International Symposium on Management of Technology, Hangzhou, P.R. China.

Hüsig, S., Hipp, C., & Dowling, M. (2005). Analysing Disruptive Potential: The Case of W-LAN and the Mobile Telecommunications Network Companies. *R&D Management, 35*(1), 17-35.

Huston, G. (2000). *Next Steps for the IP QoS Architecture*, IETF, RFC2990.

Huston, G. (2007). Quality of service-fact or fiction? Retrieved November 27, 2007, from http://www.cisco.com/web/about/ac123/ac147/ac174/ac197/about_cisco_ipj_archive_article09186a00800c8314.html

Hwang, K.-S., Tan, S.-W., Hsiao, M.-C., & Wu, C.-S. (2005). Cooperative multiagent congestion control for high-speed networks. *IEEE Transactions on Systems, Man and Cybernetics, Part B, 35*(2), 255-268.

Hyden, L.C., & Bulow, P.H. (2003). Who's talking? Drawing conclusions from focus groups: some methodological considerations. *International Journal of Social Research Methodology, 5*(4) 305-321.

Hyland, P., & Beckett, R. (2002). Learning to compete: The value of internal benchmarking. *Benchmarking: An International Journal*, *9*(3), 293-304.

Iatropoulos, A., Economides, A., & Angelou, G. (2004). Broadband investments analysis using real options methodology: A case study for Egnatia Odos S.A. *Communications and Strategies, 55*(3), 45-76.

IEEE (1995) Supplement to ISO/IEC 8802-3:1993 (1993). *Local and Metropolitan Area Networks: Media Access Control (MAC) Parameters, Physical Layer, Medium Attachment Units, and Repeaters for 100 Mb/s Operation, Type 100BASE-T (Clauses 21-30).*

IEEE 802.11. (2008). *IEEE 802.11 Standard Family*. Retrieved January 30, 2008, from http://standards.ieee.org/getieee802/802.11.html

IEEE 802.16. (2008). *IEEE 802.16 Standard Family*. Retrieved January 30, 2008, from http://standards.ieee.org/getieee802/802.16.html

IEEE 802.17 (2004). *Resilient packet ring (RPR) access method and physical layer specifications*. IEEE Standard 802.17.

IEEE 802.3 (1998). *Information Technology—Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks—Specific requirements—Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. IEEE, New York.

Ieee. (2007). IEEE802 Background. *http://www.ieee802.org/16/pub/backgrounder.html*.

Ilie, D., Popescu, A., Statistical models for Gnutella signaling traffic, Computer Networks 51 (2007) 4816-4835

Industry Canada. (2007). *Consultation on a Framework to Auction Spectrum in the 2 GHz Range including Advanced Wireless Services.*

Infoblox integrated IP address management solution. (2007). White paper, Infoblox. Retrieved April 12, 2007, from http://forms.infoblox.com/go/infoblox/wpp1001

Informa. (2006). Global Mobile: The International Business Newsletter of Mobile Communications Markets. *13*(6).

Informa. (2005). *Global Mobile Forecasts to 2010: Worldwide Market Analysis & Strategic Outlook 2005-2010* (5th ed.). London: Informa Telecoms & Media.

Informa. (2005). *Mobile Content and Services: Worldwide Market Analysis & Strategic Outlook*. London: Informa Telecoms & Media.

Informa Research (2005). *Informa research services' telecom market segmentation and customer profiling program*. **Accessed in March 2006** http://www.markettrends.com/industry/telecom/telecomproj.htm#5

Innes, J., & Mitchell, F. (1991). *Activity based cost management: A case study of development and implementation*. London: CIMA.

Innes, J., & Mitchell, F. (1995). A survey of activity-based costing in the UK's largest companies. *Management Accounting Research*, *6*, 137-153.

Innes, J., & Mitchell, F. (1996). A review of activity-based costing practice. In Drury, C. (Ed.) *Management accounting handbook*, 2nd edition. CIMA Publishing.

Innes, J., Mitchell, F., & Sinclair, D. (2000). Activity-based costing in the UK's largest companies: A comparison of 1994 and 1999 survey results. *Management Accounting Research*, *11*, 349-362.

Inoue, A., Takahashi, S., & Nishimatsu, K. (2003). Service Demand Analysis Using Multi-Attribute Learning Mechanisms. *KIMAS*, IEEE, (pp. 634-639).

Insight Reports (2005). *Telecommunications and VoIP in Asia, Oceania and the mideast: a market perspective on the major economies 2005-2011.* The insight research corporation, USA.

Intel (2000). *Caching for Improved Content Delivery.*

Intel Solutions. (2005). *Digital Community Best Practices* (White Paper).

Intel. (2005, August). *Core Technologies for Developing a Digital Community Framework Wireless Fabric, GIS Portal, and Government Federated Services Bus Technologies* (White Paper)

International Telecom Union (2002). *Asia-Pacific Telecommunications Indicators.* Geneva: ITU.

International Telecom Union (2001). *Mobile Virtual Network Operators.* Geneva. Geneva: ITU.

International Telecom Union (2004*). Trends in telecommunication reform: Effective regulation*. Geneva: ITU.

International Telecommunication Union (1992). *Management Framework for Open Systems Interconnection (OSI) for CCITT Applications*. Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.700-199209-I/en

International Telecommunication Union (1997). *Information Technology – Open Systems Interconnection – Common Management Information Protocol: Specification.* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.711-199710-I/en

International Telecommunication Union (2000). *Series M: TMN and Network Maintenance: International Transmission Systems, Telephone Circuits, Telegraphy, Facsimile, and Leased Circuits.* Retrieved January 2008 from http://www.itu.int/rec/t-rec-m.3400-200002-I/en

International Telecommunication Union (2002). *Information Technology – Abstract Syntax Notation One (ASN.1): Specification of Basic Notation.* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.680/en

International Telecommunication Union (2002). *Information Technology –ASN.1 Encoding Rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER).* Retrieved January 2008 from http://www.itu.int/rec/T-REC-X.690/en

International Telecommunications Union. (1996). Methods for subjective determination of transmission quality [Electronic Version]. Retrieved October 29, 2007 from http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.800-199608-I!!PDF-E&type=items.

Intime, Inc. (2007). http://www.gointime.com/ (Verified January 23, 2008)

IP address management demands multi-core architectures. (2006). White paper, BlueCat Networks. Retrieved June 16, 2007, from http://www.bluecatnetworks.com/demo/Proteus_Multi-Core_Architecture.pdf

Irani, Z., Beskese, A., & Love, P. E. (2004). Total quality management and corporate culture: constructs of organisational excellence. *Technovation, 24,* 643–650.

Irani, Z., Sharp, J. M., & Kagioglou, M. (1997). Improving business performance through developing a corporate culture. *The TQM Magazine, 9*(3), 206–216.

Iraqi, Y., & Boutaba, R. (2000). Resource management issues in future wireless multimedia networks. *Journal of High Speed Networks*, *9*, 231-260.

Irvine, C. (2001). Emerging Value Propositions for M-Commerce. *Journal of Business Strategies, 18(2),* 133-149.

Isabella, L. A. (1990). Evolving interpretations as a change unfolds: how managers construe key organizational events. *Academy of Management Journal, 33*(1), 7-41.

Ishara, K. S. (2003). *Mobile Commerce- Usage and Usability.*

ISO (1998) *ISO 9241 – 11, Ergonomics Requirements for office work with visual display terminals (VDTS), Part 11: Guidance on Usability.*

ISO/IEC 9596 (1991). *Common Management Information Protocol Specification*.

Itami, H. (1987). *Mobilizing Invisible Assets.* Cambridge, MA: Harvard University Press.

ITU (2008). *3G network standards.* Retrieved January 30, 2008, from http://www.itu.int/ITU-T/

ITU-T (2004). *Asia-Pacific Telecommunication Indicators 2004.*

ITU-T (2005). Recommendation G.8110. *T-MPLS Layer Network Architecture*. January 2005.

ITU-T (2000). *Telecommunication Reform: Interconnection Regulation.*

ITU-T Rec. H.323. (1998). *Packet-Based Multimedia Communications Systems.* Geneva.

ITU-T SG3 (2003). *Report of The Joint Rapporteurs' Group Meeting Held in Brussels* on 20-21 Oct. 2003-Final version 5 Nov. 03-Rev.1," TU-T SG3 document, COM3-D53 Rev.2-E, Nov.

Ivancevich, J. M. (2001). *International Human Resource Management*, 8th ed. New York, NY: McGraw-Hill/Irwin.

Jackson, L., Rouskas, G., & Stallman, M. (2007). The directional p-median problem: definition, complexity, and algorithms. *European Journal of Operational Reseach*, *179*, 1097-1108.

Jackson, T. (2002). *Cross-cultural Approach*. London: Sage Publications.

Jacobs, G., & Rodgers, C. (1998). ISDN-based distance learning: Tangible costs, intangible benefits. *Computers and Education*, *31*(1), 41-53.

Jacome, R., Lisboa, J., & Yasin, M. (2002). Time-based differentiation – An old strategic hat or an effective strategic choice? An empirical investigation, *European Business Review, 14*(3), 184-93.

Jafar, S. A., & Srinivasa, S. (2007). Capacity limits of cognitive radio with distributed and dynamic spectral activity. *IEEE Journal on Selected Areas in Communications*, *25*(3), 529 – 537.

Jakobson, G., & Weissman, M. D. (1993). Alarm Correlation. *IEEE Network* (pp. 52-59).

Jamalipour, A., Wada, T., & Yamazato, T. (2005). A tutorial on multiple access technologies for beyond 3G mobile networks. *IEEE Communications Magazine*, *43*(2), 110-117.

Janakiraman, R., Waldvogel, M., & Qi Z. (2003). Indra: a peer-to-peer approach to network intrusion detection and prevention, *Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, (pp. 226 -231). Dept. of Comput. Sci. & Eng., Washington Univ., St. Louis, WA, USA.

Jang, B. (2003). Vertical structure and network competition in telecommunications services market. *Information Society Research*, Fall: 81-101.

Jauch, L., Osborn, R., & Martin, T. (1980). Structured content analysis of cases: A complementary method of organizational research. *Academy of Management Review*, *5*, 517-526.

Java Call Control (JCC) API (2001). Version 1.0, *Overview of the API*. Java Technology. The JAIN initiative, at http://java.sun.com/products/jain/

Javalgi, R. R. G., Griffith, D. A., & White, D. S. (2003). An empirical examination of factors influencing the internationalization of service firms. *Journal of Services Marketing*, *17*(2), 185-201.

Jeffres, L., Atkin, D., Neuendorf, K., & Lin, C. (2004). The influence of expanding media menus on audience content selection. *Telematics & Informatics, 21*, 317-334.

Jelger, C., Noel, T., & Fray, A. (2004). *Gateway and address auto-configuration for IPv6 ad-hoc networks*, IETF draft, April 2004, draft-jelger-manet-gateway-autoconf-v6-02.txt. Retrieved December 14, 2007, from http://tools.ietf.org/id/draft-jelger-manet-gateway-autoconf-v6-02.

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of ACM*, *44(4)*, 35-41.

Jennings, N. R., Parsons, S., Noriega, P., & Sierra, C. (1998). On argumentation Based Negotiation. *Proceedings of the Int. Workshop on Multi-Agent Systems.* (pp. 1-7). MIT.

Jensen, F. (2001a). *Bayesian Networks and decision graphs*. New York, USA: Springer-Verlag.

Jensen, F. V., Kjærulff, U., Kristiansen, B., Langseth, H., Skaanning, C., Vomlel, J., & Vomlelová, M. (2001b). SACSO methodology for troubleshooting complex systems. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 5(4), 321-333.

Jeong, H., Kim, D., Park, J., & Toh, C. (2007). *Passive duplicate address detection for on-demand routing protocols*. IETF draft, April 2007, draft-jeong-autoconf-pdad-on-demand-01.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-jeong-autoconf-pdad-on-demand-01

Jeong, H.-Y., & Seo, S.-W. (2005). A binary (0-1) linear program formulation for the placement of limited-range wavelength converters in wavelength-routed WDM networks. *Journal of Lihtwave Technology, 23*(10), 3076-3086.

Jeong, J., Park, J., Kim, H., Jeong, H., & Kim, D. (2006). *Ad-hoc IP address auto-configuration*. IETF draft, January 2006, draft-jeong-adhoc-ip-addr-autoconf-06.txt. Retrieved December 14, 2007, from http://ietfreport.isoc.org/idref/draft-jeong-adhoc-ip-addr-autoconf/.

Jho, W. (2007). Liberalisation as a development strategy: network governance in the Korean mobile telecom market. *Governance: An International Journal of Policy, Administration, and Institutions, 20*(4), 633-654.

Jia, X., Hu, X.-D., & Du, D.-Z. (2002). *Multiwavelength optical networks.* Kluwer Academic Publishers.

Jiang, W., & Schulzrinne, H. (2000). Modeling of packet loss and delay and their effect on real-time multimedia service quality. *In Proceedings of International Workshop* *on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2000).*

Jinpei, C. (2002). *China research and development goes into orbit.* http://www.my-esm.com/article

Johanson, J., & Vahlne, J.-E. (1977). The Internationalization Process of the Firm - a Model of Knowledge Development and Increasing Foreign Market Commitments. *Journal of International Business Studies*, *8*(1), 23-32.

Johanson, J., & Wiedersheim-Paul, F. (1975). The Internationalization of the Firm: Four Swedish Cases. *Journal of Management Studies, 12*(3), 305-322.

Johnson, D. M. (2004). Adaptation of organizational change models to the implementation of quality standard requirements, *International Journal of Quality & Reliability Management*, *21*(2), 154-174.

Johnson, H. T., & Kaplan, R. S. (1991). *Relevance lost: The rise and fall of management accounting*, 2nd edition. Boston: Harvard Business School Press.

Johnson, J. L. (2004). *Fundamentals of Cellular Network Planning and Optimisation*. West Sussex, England: Wiley.

Johnston, A., Donovan, S., Sparks, R., Cunningham, C., & Summers, K. (2003). *Session Initiation Protocol (SIP) Basic Call Flow Examples*. Request for Comments (Best Current Practice), 3665, Internet Engineering Task Force.

Johnston, N., & Aghvami, H. (2007). Comparing WiMAX and HSPA : a guide to the technology. *BT Technology Journal, 25*(2), 191-199.

Jones, M. A., Mothersbaugh, D. L., & Betty, S. E. (2000). Switching barriers and repurchase intentions in services. *Journal of Retailing*, *76*(2), 259-272.

Jones, M., & Marsden, G. (2006). *Mobile Interaction Design*. Chichester: John Wiley & Sons.

Jones, N. B., Herschel, R. T., & Moesel, D. D. (2003). Using knowledge champions to facilitate knowledge management. *Journal of Knowledge Management*, *7*(1), 49-63

Jones, S., & Salsas, P. (2006). E-Communications: Investment and the Regulatory Framework. *Communications & Strategies*, *64*, 29-39.

Jones, T. C., & Dugdale, D. (2002). The ABC bandwagon and the juggernaut of modernity. *Accounting, Organizations and Society*, *27*, 121-163.

Jong, Y. C., Markus, J., & Susanne, W. (2005). Balancing auditability and privacy in vehicular networks. In Y. Xiaos (Ed.), *The First ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks* (pp. 79-87). ACM: New York, USA.

Jorgenson, D. (2001). Information technology and the U.S. economy. *American Economic Review, 91*, 1-32.

Joskow, P., & Rose, N. (1989). The effects of economic regulation. In R. Schmalensee and R. Willig (Eds.), *Handbook of Industrial Organization, 2*, 1449-1506. Amsterdam: North-Holland.

Jost, J. (2004). *Can MVNOs succeed in emerging markets?* Pyramid Research.

Jovicic, A., & Viswanath, P. (2006). Cognitive Radio: An information-theoretic perspective. *IEEE International Symposium on Information Theory, 1*, 2413-2417.

JP Morgan (2006). *The fibre battle*.

JSR 32 (2006). *JAIN SIP API Specification*. http://jcp.org/en/jsr/detail?id=32

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology, 50*, 359-94.

Jung, I., Gayle, P., & Lehman, D. (2008). Competition and investment in telecommunications. *Applied Economics*, *40*, 303-313.

Jurca, R., & Faltings, B. (2005). Reputation-based Service Level Agreements for Web services. In *Proceedings of Third International Conference on Service Oriented Computing* (ICSOC 2005), Amsterdam, The Netherlands.

Kaashoek, M. F., & Karger, D. R. (2003). Koorde: A simple degree-optimal distributed hash table. In *Proceedings of* *the 2nd International Workshop on Peer-to-Peer Systems (IPTPS03)*. Berkeley, CA.

Kahn, A., Tardiff, T., & Weisman, D. L. (1999). The Telecommunications Act at three years: An economic evaluation of its implementation by the Federal Communications Commission. *Information Economics and Policy, 11*(4), 319-365.

Kahn, J. (1998). The World's most admired companies. *Fortune of 26 October*, 206.

Kakihara, M., & Sørensen, C. (2002). Mobility. *35th Annual Hawaii International Conference on System Sciences.* Washington, DC: IEEE Computer Society.

Kalsom, S., & Syed Noh, S. A. (June, 2006). *KM strategy for E-Government: An exploratory study of local authorities in Malaysia*. Paper presented at the Knowledge Management International Conference and Exhibition 2006 (257-265), Legend Hotel, Kuala Lumpur, Malaysia.

Kanellopoulos, D., Pintelas, P., & Giannoulis, S. (2006). QoS in wireless multimedia networks. *Annals of Mathematics, Computing & TeleInformatics*, *1*(4), 66-75.

Kannan, P., Chang, A., & Whinston, A. (2001). Wireless Commerce: Marketing Issues and Possibilities. *34th Hawaii International Conference on System Sciences.* Washington, DC: IEEE Computer Society.

Kansal, A., & Karandikar, A. (2001). *Jitter-free audio playout over best effort packet networks.* ATM Forum International Symposium, New Delhi, India.

Kaplan, R. S. (1984). Yesterday's accounting undermines production. *Harvard Business Review*, July-August, 95-101.

Kaplan. R. S., & Norton, D. P. (1992). The balanced scored measures that drive performance. *Harvard Business Review*, *70*(1), 71-79.

Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Karahanna, E., Evaristo, R., & Srite, M. (2005). Levels of culture and individual behavior: An integrative perspective. *Journal of Global Information Management, 13*(2), 1–20.

Karathanos, P. (1998). Crafting corporate meaning (developing corporate culture). *Management Decision, 36*(2), 123–132.

Karlson, B., Bri, A., Link, J., Lönnqvist, P., & Norling, C. (2003). *Wireless Foresight: Scenarios of the Mobile World in 2015.* John Wiley and Sons.

Karol, M., Krishnan, P., & Li, J. J. (2005). VoIP protection and performance Improvement. *Telecommunication Systems*, *28*(3-4), 351.

Karpoff, J. M., & Lott, J. R. (1999). On the determinants and importance of punitive damage awards. *Journal of Law and Economics, 42*, 527-573.

Kates, S. M., & Robertson, J. (2004). Adapting action research to marketing. A dialogic argument between theory and practice. *European Journal of Marketing*, *38*(¾),418-432.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition and compatibility. *American Economic Review*, *75*(3), 424-440.

Katzela, I., & Schwartz, M. (1995). Schemes for fault identification in communication networks. *IEEE/ACM Transactions on Networking*, *3*(6), 753-764.

Kautz, K., & Larsen, E. A. (2000). Diffusion Theory and Practice: Disseminating quality management and software process improvement innovations. *Information Technology and People, 13*(1), 11-26.

Kawsar, F., Fujinami, K., Pirttikangas, S., & Nakajima, T. (2006). Personalization and Context Aware Services: A Middleware Perspective. *2nd International Workshop on Personalized Context Modeling and Management for UbiComp Applications.* Berlin: Springer-Verlag.

Kay, S. M. (1). (1993). *Fundamentals of Statistical Signal Processing*: *Estimation Theory*, Prentice Hall Signal Processing Series (pp. 1-217).

KDDI (2003). *A Suggestion for The Amendment of Recommendation D.50.* ITU-T SG3 document COM 3-D47-E, June.

Kearney, A. T. (2001). *Satisfying the experienced online shopper: Global e-shopping survey*. A.T. Kearney Ltd., Inc. London.

Keen, P. (1997). *Business multimedia explained: A manager's guide to key terms & concepts*. Cambridge, MA: Harvard Business School Press.

Kelaskar, M., Matossian, V., Mehra, P., Paul, D., & Parashar, M. (2002). A study of discovery mechanisms for peer-to-peer applications. In *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid Workshop on Global and Peer-to-Peer on Large Scale Distributed Systems* (pp. 444-445). Berlin, Germany.

Keller, A., & Ludwig, H. (2003). The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web services. *Journal of Network and Systems Management, Special Issue on ``E-Business Management*, (pp. 27-33).

Kempken, S. and Luther, W. Modeling of H.264 high definition video traffic using discrete time semi-Markov processes, Proc. 20. International Teletraffic Congress, Ottawa, Canada, Springer, LNCS 4516 (2007) 42-53

Kende M. (2000). *The Digital Handshake: Connecting Internet Backbones*. Federal Communications Commission, Office of Plans and Policy, Working Paper no. 32, Washington DC. Sep.

Keng, S., & Ee-Peng, L. (2001). Mobile Commerce: Promises, Challenges, and Research Agenda. *Journal of Database Management, 12(3)*, 4-24.

Kennington, J. L., Olinick, E. V., & Spiride, G. (2007). Basic mathematical programming models for capacity allocation in mesh-based survivable networks. *Omega*, *35*, 629-644.

Kennington, J., Olinick, E., Ortyinski, A., & Spiride, G. (2003). Wavelength routing and assignment in a survivable WDM mesh network. *Operations Research*, *51*(1), 67-79.

Kephart, J., & Chess, D. (2003). The vision of autonomic computing. *Computer, 36*(1), (pp. 41-50).

Kerravala, Z. (2005). *IPv6 success depends on the proper management tools and process*. Enterprise Computing & Networking, Yankee Group Consulting Report. Retrieved September 3, 2007, from http://www.networkworldpartners.com/lucent/pdf/Yankee_IPv6.pdf

Keryer, T.N. (2001). I-mode: a successful launch of the mobile Internet market. *Alcatel Telecommunications Review*, (pp. 64-67).

Kesh, S., & Raja, M. K. (2005). Development of a qualitative reasoning model for financial forecasting. *Information Management & Computer Security, 13*(2), 167-179.

Kevin, H., David, Z., & Cristina, R. (2007). *A Survey of attack and defense techniques for reputation systems.* From http://www.cs.purdue.edu/homes/zagedj/docs/reputation_survey.pdf

Khalil, I., Ibrahim, R., & Kronsteiner, G. K. (2004). Agent-Based Mobile Auctions: The Flea Market Scenario. *Proceedings of 2004 International Research Conference on Innovations in Information Technology (IIT2004),* Dubai.

Khanafer, R., Moltsen, L., Dubreil, H., Altman, Z., & Barco, R. (2006). A Bayesian approach for automated troubleshooting for UMTS networks. *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06).* Karlsruhe, Germany.

Khanafer, R., Solana, B., Triola, J., Barco, R., Moltsen, L., Altman, Z., & Lázaro, P. (2008). Automated diagnosis for UMTS networks using Bayesian Network Approach. *IEEE Transactions on Vehicular Technology*. In press.

Khoshgoftaar, T. M. & Abushadi, M. E. (2004). Resource-sensitive intrusion detection models for network traffic. *Eighth IEEE International Symposium on High Assurance Systems Engineering,* (pp. 249 – 258). Department of Computer Science & Engineering, Florida Atlantic University, Boca Raton, FL, USA.

Kielmann, T., & Hofman, R., (1999). *The Genetic Algorithms* homepage: ftp.cs.vu.nl/pub/papers/ai/GA/. E-mail: kielmann@cs.vu.nl.

Kijl, B., Bouwman, H., Haaker, T., & Faber, E. (2005). *Dynamic Business Models in Mobile Service Value Networks: A Theoretical and Conceptual Framework (FRUX/D2.2-II).* The Netherlands: TU Delft.

Killström, U., Galli, L., Haaker, T., Immonen, O., Kijl, B., Pitkänen, O., Saarinen, P. J., & Virola, H. (2006). *Initial marketplace dynamics (incl. business models) analysis (IST-2004-511607 MobiLife).* Finland: Information Society Technologies.

Kim J. (1998). Hierachical structure of intranet functions and their relative importance: Using the analytic hierarchy process for virtual organizations. *Decision Support Systems*, *23*, 59-74.

Kim, C.-H (2005). *Trend of Digital Home Network Development,* Korea Electronics Technology Institute.

Kim, D., Cavusgil, S. T., & Calantone, R. J. (2006),. nformation system innovations and supply chain management: channel relationships and firm performance. *Journal of the Academy of Marketing Science*, *34*(1), 40-54.

Kim, J. (2001). Analyzing mobile Internet users. *Presented at the CHI 2001 Workshop on mobile communications.*

Kim, M. K., & Park, J. H. (2008). Characteristics of subscriber and usage for WiBro and HSDPA at the introductory market in Korea. *Korean Radio Promotion Association,* Arpil-March.

Kim, M. K., Lee, M. H., & Park, J. H. (2005) Impacts of the introduction of mobile number portability on customer switching intention. *Telecommunications Review*, *15*(4), 667-679.

Kim, M. K., Park, M. C., & Jeong, D. H. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in the Korean mobile telecommunication services. *Telecommunications Policy*, *28*(2), 145-159.

Kim, M. K., Park, M. C., Jeong, D. H., & Park, J. H. (2003). An empirical study on the adjustment Effect of switching barriers on customer retention in mobile telecommunications services. *The Journal of MIS Research*, *13*(3), 107-130.

Kim, M., Park, M., & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunications services. *Telecommunications Policy, 28*, 145–159.

Kim, S.-W., Seo, S.-S., & Kim, S. (2000). Regenerator placement algorithms for connection establishment in all-optical networks. *GLOBECOM - Global Telecommunications Conference IEEE,* (pp. 1205-1209).

King, J. L., Gurbaxani, V., Kraemer, K. L., McFarlan, K. L., Raman, K. S., & Yap, C. S. (1994). Institutional factors in information technology innovation. *Information Systems Research*, *5*(2), 139-169.

King, R. (2007). Line goes dead for a Web caller. *Business Week Online*, 3-3.

Kirakowski, J., & Corbett, M. (1988). Measuring User Satisfaction. In D. M. Jones, and R. Winder (Eds.), *People and Computers, IV.* Cambridge, UK: Cambridge University Press.

Kirkpatrick, S., Gerlatt, C. D. Jr., & Vecchi, M.P. (1983). Optimization by Simulated Annealing. *Science*, *220*, 671-680.

Kirpalani, M., & Luostarinen, R. (1999, Dec 14). *Dynamics of Success of SMOPEC Firms in Global Markets.* Paper presented at the EIBA Annual Conference, Manchester.

Kirstädter, A., Gruber, C., Riedl, J., & Bauschert, T. (2006). Carrier-grade Ethernet for packet core networks. *Proc., SPIE,* 6354. 635414-1–635414-14, Gwangju, Korea, Sept. 2006.

Klemperer, P. (2002). How (not) to run auctions: The European 3G telecom auctions, *European Economic Review*, *46*(4-5), 829-845.

Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research, 53*, 25-37.

Knight, G. A. (1999). International services marketing: review of research 1980-1998. *Journal of Services Marketing, 13*(4/5), 347-360.

Knight, G. A., & Cavusgil, S. T. (1996). The Born Global firm: a challenge to traditional internationalization theory. *Advances in International Marketing, 8*, 11-26.

Knights, D., & Willmott, H. (1995). Culture and control in a life insurance company. *Studies in Cultures, Organizations and Societies, 1*,(1), 29-46.

Kogut, B., & Singh, H. (1986). Entering the United States by acquisition or joint venture, country patterns and cultural characteristics, Working Paper, Reginald H. Jones Center, Wharton School, University of Pennsylvania, Philadelphia, PA.

Koku, P. S., Qureshi, A. A., & Akhigbe, A. (2001). The effects of news on initial corporate lawsuits. *Journal of Business Research*, *53*, 49-55.

Kolbe, A. L., & Tye, W. B. (1990). The Duquesne opinion: How much "Hope" is there for investors in regulated firms? *Yale Journal on Regulation, 8*(1), 113-157.

Kollmann, T. (2000). Competitive strategies for electronic marketplaces. *Electronic Markets*, *10*(2), 102–109.

Kollmann, T. (2001). Measuring the acceptance of electronic marketplaces. *Journal of Computer Mediated Communication*, *2*(6).

Kollmann, T. (1998). Marketing for electronic market places – the relevance of two critical points of success". *Electronic Markets*, *8*(4), S. 36–39.

Kollmann, T. (2005). The matching function for electronic market places – determining the probability of coordinating of supply and demand. *International Journal of Electronic Business*, *5*(3), 461–472.

Kollmann, T. (1999). Virtual marketplaces: Building management information systems for internet brokerage. *Virtual Reality*, *4*(4), 275–290.

Kollmann, T. (2006). What is e-entrepreneurship? – Fundamentals of company founding in the net economy. *International Journal of Technology Management*, *33*(4), 322–340.

Kong, X., & Mase, K. (2003). Dynamic routing with endpoint admission control for VoIP networks. *IEEE International Conference on Communications, 2003. ICC '03, 3*, 1728-1732.

König, H. (2003). *Protocol Engineering* (in German). Weisbaden, B. G. Teubner Verlag.

Konorski, J. (2006). A game-theoretic study of CSMA/CA under a backoff attack. *IEEE/ACM Transactions on Networking, 14*(6), 1167-1178.

Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities, in ed. Koopmans. *Analysis of Production and Allocation* (33-97). New York, NY: Wiley.

Korea Home Network Industries Association (2005). In M. & B. Jinhan (Eds.), *Status of Local and Overseas Home Network Industry and Future Development Strategy.*

Korea Home Network Industries Association (2005). *Status of Local and Overseas Home Network Industry and Future Development Strategy.* Jinhan M & B.

Kotter, J. P., & Heskett, J. L. (1992). *Corporate Culture and Performance.* New York: The Free Press.

Koukia, S., Rigou, M., & Sirmakessis, S. (2006).The Role of Context in m-Commerce and the Personalization Dimension. *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.* Washington, DC: IEEE Computer Society.

Kraemer, K., & Dedrick, J. (2002). Dell computer: Organization of a global production network." Available at: http://www.crito.uci.edu/GIT/publications/pdf/dell.pdf Retrieved: 6 October 2006.

Krafft, J. (2003). Vertical structure of the industry and competition: an analysis of the evolution of the info-communications industry. *Telecommunications Policy, 27*(8/9), 625-649.

Kraus, S. (1992). Negotiation in a non-cooperative environment. *Journal of Experimental and Theoretical Artificial Intelligence*, 3(4), 255-281.

Krausman, A. S., & Nussbaum, M. A. (2007). Effects of wearing chemical protective clothing on text entry when using wearable input devices. *International Journal of Industrial Ergonomics, 37*(6), 525-530.

Kretkowski, P. D. (2006). *Top 10 VoIP predictions for 2007.* Retrieved October 5, 2007, from http://www.voip-news.com/feature/top-ten-voip-predictions-122706/

Kridel, D., Sappington, D. E. M., & Weisman, D. L. (1996). The effects of incentive regulation in the telecommunications industry: A survey. *Journal of Regulatory Economics, 9*(3), 269-306.

Krishna, S., Leping, H., Mingyan, L., Radha, P., Matsuura, K., & Sezaki, K. (2005). CARAVAN: Providing location privacy for VANET. In A. Amer (Ed.), *The Embedded Security in Cars* (pp. 367-383). ACM: Germany.

Kristensson, S. (2001). *Mobile Virtual Network Operators. Assessing MVNO business opportunities.* School of Business, University of Stockholm, Master of Science Thesis, June 2001.

Kristof, C., & Dirk, V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, *Expert Systems with Applications*, *34*(1), 313-327.

Kronsteiner, R., Khalil, I., & Ibrahim, G. K. (2005). Agents for Mobility Engineering in m-Business. *ACM International Conference Proceeding Series; Proceedings of the 4th international symposium on Information and communication technologies, session Information technology, 92,* 57-62. Trinity College Dublin.

Krupinski, D. (1997), *Introduction to Internet Telephony*. http://www.phonezone.com/tutorial/it-intro.htm.

Kubat, P., & Macgregor Smith, J. (2001). A multi-period network design problem for cellular telecommunications systems. *European Journal of Operational Research*, 439-456.

Kuhn, D. R. (1997). Sources of failure in the public switched telephone network. *Computer, 30*(4), 31-36.

Kumar, R., & Das, T. K. (2007). Interpartner legitimacy in the Alliance Development Process. *Journal of Management Studies,* 1-29.

Kumar, S. (2004). Mobile communications: global trends in the 21st century. *International Journal of Mobile Communications, 2*(1), 67-86.

Kurkovsky, S. (2007). Pervasive Computing: Past, Present and Future. *5th International Conference on Information and Communication Technology*. Washington, DC: IEEE Computer Society.

Kurkovsky, S., & Harihar, K. (2006). Using Ubiquitous Computing in Interactive Mobile Marketing. *Personal and Ubiquitous Computing, 10*(4). Berlin: Springer-Verlag.

Kurkovsky, S., Strimple, D., Nuzzi, E., & Verdecchia, K. (2008). Mobile Voice Access in Social Networking Systems. *5th International Conference on Information Technology: New Generations*. Washington, DC: IEEE Computer Society.

Laanti, R. (2008). *Internationalisation Strategies of Companies in Service Industries: A Study of National Telecommunication Operators from Small and Open Economies (SMOPECs).* Unpublished PhD, The University of Adelaide, Adelaide.

Laanti, R., Gabrielsson, M., & Gabrielsson, P. (2007). The globalization strategies of business-to-business born global firms in the wireless technology industry. *Industrial Marketing Management, 36*(8), 1104-1117.

Laanti, R., McDougall, F., & Baume, G. (2006). *Stages Or Leaps? - Market Strategies Of Telecommunication Operators From Small And Medium Sized Countries.* Paper presented at the 48th Annual Meeting of the Academy of International Business, Beijing.

Lacatus, C., & Popescu, D. C. (2007). Adaptive Interference Avoidance for Dynamic Wireless Systems: A Game Theoretic Approach. *IEEE Journal of Selected Topics in Signal Processing, 1*(1), 189-202.

Laffont, J. J., & Tirole, J. (2001). *Competition in Telecommunications.* MIT Press.

Laffont, J. J., Rey, P., & Tirole, J. (1997). Competition between telecommunications operators. *European Economic Review, 41*, 701-711.

Lai, S., Trueblood, P., & Wong, K. (1999). Software Selection: a case study of the application of the analytic hierarchy process to the selection of multimedia authoring system. *Information & Management, 36*, 221-232.

Laiho J., Wacker, A., & Novosad, T. (2002a). *Radio network planning and optimisation for UMTS*. Chichester, UK: Wiley

Laiho, J., Kylväjä, V., & Höglund, A. (2002). Utilisation of Advanced Analysis Methods in UMTS Networks. *Proceedings of the IEEE Vehicular Technology Conference VTC'02* (pp. 726-730). Birmingham, USA

Laiho, J., Ravio, K., Lehtimäki, P., Hätönen, K., & Simula, O. (2005). Advanced Analysis Methods for 3G Cellular Networks. *IEEE Transactions on Wireless Communications*, *4*(3), 930-942.

Lange, J. (1994). ISDN videoconferencing for education and training. In R. Mason, Bacsich, P. (Ed.), *ISDN Applications in Education and Training* (pp. 127-143). London: IEEE Press.

Lapid, Y., Rom, R., & Slidi, M. (1998). Analysis of discarding policies in high-speed networks. *IEEE Journal on Selected Areas in Communications*, *16*(5), 764-777.

Laroia, R., Uppala, S., & Li, J. (2004). Designing a mobile broadband wireless access network. *IEEE Signal Processing Magazine*, *21*(5), 20- 28.

Laughton, D. (1991). The Valuation of Offshore Oilfield Development In Leases. Advances In M. Breton &G. Zaccour (Eds.),Operations Research in the Oil and Gas Industry and Editions Technip (pp.63-82), Paris 1991.

Laukkanen, T. (2007). Internet vs. Mobile Banking: comparing customer value perceptions. *Business Process Management Journal*, *13*(6), 788-797.

Laurence, M. (2005). How echo affects VOIP. *Business Communications Review*, *35*(8), 50.

Lauther, U., Winter, T., & Ziegelmann, M. (2003). Proximity Graph based Clustering Algorithms for Optimized Planning of UMTS Access Network Topologies. *IEEE International Conference on Telecommunications (ICT 2003)*, 1329-1334.

Lázaro, P., Barco, R., & Hermoso, J. M. (2002) Diagnosis of Earth Stations Using Bayesian Networks. *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*(pp. 268-272). Málaga, Spain.

Lee J., Lee, J., & Freick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link; Mobile phone service in france. *Journal of services Marketing*, *15*(1), 35-48.

Lee, C. H. (2001). *State regulatory commission treatment of advanced services: Results of a survey*. Columbus, OH: National Regulatory Research Institute.

Lee, H. G., & Clark, T. H. (1996). Impacts of the electronic marketplace on transaction cost and market Structure. *International Journal of Electronic Commerce*, *1*(1), 127–149.

Lee, H. S., & Kim, S. Y. (2003). A study on the moderate effect of switching cost and the customer satisfaction affecting customer loyalty. *Telecommunications Review*, *13*(6), 1027-1039.

Lee, I. (2005a). An analytical model of e-recruiting investment decision: An economic employment approach. *IEEE Transactions on Engineering Management, 52*(4), 486-496.

Lee, I. (2005b). The evolution of e-recruiting: A content analysis of fortune 100 career web sites. *Journal of Electronic Commerce in Organizations*, *3*(3), 57-68.

Lee, J. S, Kim, Y. B., Lee, J. D., & Park, Y. R. (2006) Estimating the Switching Cost and its Implications for Market Structure: Case of the Korean Mobile Communications Market and Number Portability. *The Korean Journal of Industrial Organization*, *12*(2), 33-58.

Lee, Y. (2004). Customer service and organizational learning in the context of strategic marketing. *Marketing Intelligence & Planning*, *22*(6), 652–662

Lee, Y. E., & Benbasat, I. (2003). Interface design for mobile commerce. *Communications of the ACM*, *46*(12), 49-52.

Lehr, W. (2002). Vertical integration, Industry Structure, and Internet Telephony. In McKnight, Lehr, and Clark (Ed.), *Internet Telephony.* The MIT Press.

Lehr, W., & McKnight, L. W. (2003). Wireless internet access: 3G vs. WiFi? *Telecommunications Policy, 27*(5-6), 351-370.

Lehtimäki, P., & Raivio, K. (2005a). A SOM Based Approach for Visualization of GSM Network Performance Data. *Proceedings of the International Symposium on Intelligent Data Analysis.* Madrid, Spain.

Lehtimäki, P., & Raivio, K. (2005b). A Knowledge-Based Model for Analyzing GSM Network Performance. *Proceedings of the International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*. Bari, Italy.

Leinmuller, T., Schoch, E., Kargl, F., & Maihofer, C. (2005). Influence of falsified position data on geographic ad-hoc routing. In R. Molva, G. Tsudik, & D. Westhoff (Eds.), *The second European Workshop on Security and Privacy in Ad hoc and Sensor Networks* (pp. 102-112). Springer.

Leland, W., Taqqu, M.S., Willinger, W. and Wilson, D.: On the self-similar nature of Ethernet traffic, IEEE/ACM Trans. on Networking 2 (1994) 1-15

Lemos, A. D., & Porto, A. C. (1998). Technological forecasting techniques and competitive intelligence: tools for improving the innovation process. *Journal of Industrial Management & Data Systems, 98*(7), 330 – 337.

Lempiäinen, J., & Manninen, M. (2001). *Radio interface system planning for GSM/GPRS/UMTS*. Kluwer Academic Publishers. Dordrecht, The Netherlands.

Leong, B., & Li, J. (2004). Achieving one-hop DHT lookup and strong stabilisation by passing tokens. In *Proceedings of the 12th International Conference on Networks (ICON2004).* Singapore.

Leske, N. (2008). Mobile ads take baby steps to reach consumers. *Reuters Technology news*, Retrieved on February 16, 2008 from http://www.reuters.com/article/technologyNews/idUSL1445372720080214

Leslie, C. (2007). Courts rebuff most of Vonage's patent arguments. *USA Today*, Retrieved November 10, 2007 from Academic Search Premier EBSCO database.

Lessiter, F. (2006). Auto-Steer's New Frontier – Precise implement steering. *Farm Eguipment*. http://www.gps-farm.com/_resources/uploads/files/Farm%20

Leung, L., & Wei, R. (2000). More than just talk on the move: Uses and gratifications of the cellular phone. *J&MC Quarterly , 77*, 308-320.

Levinson, M. (2005). The ABCs of KM. *Knowledge Management Research Centre*. Retrieved January 22, 2008, from http://www.cio.com/research/knowledge/edit/kmabcs.html.

Levis, P., Asgari, H., & Trimintzios, P. (2004). Consideration on Inter-domain QoS and Traffic Engineering issues Through a Utopian Approach. *SAPIR-2004 workshop of ICT-2004*, (C) Springer-Verlag.

Levitt, T. (1983). The Globalization of Markets. *Harvard Business Review, 61*(3), 92-101.

Levy, R., Nagarajarao, J., Pacifici, G., Spreitzer, M., Tantawi, A., & Youssef, A. (2003). Performance Management for Cluster Based Web services. In *Proceedings of the 8th IFIP/IEEE International. Symposium on Integrated Network Management (IM2003).*

Li, F., & Whalley, J. (2002). Deconstruction of the Telecommunication Industry: From Value Chains to Value Networks. *Telecommunications Policy, 26*(9/10), 451-472.

Li, F., & Whalley, J. (2004). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy*, *26*(9), 451-472.

Li, H., Bar-Ness, Y., Abdi, A., Somekh, O. S., & Su, W. (2006). OFDM modulation classification and parameters extraction. *1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 1*, 1-6.

Li, T., Mow, W. H., Lau, V. K. N., Siu, M., Cheng, R. S., & Murch, R. D. (2007). Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference. *IEEE Journal on Selected Areas in Communications, 25*(3), 566-575.

Li, W., & McQueen, J. (2008). Barriers to mobile commerce adoption: an analysis framework for a country-level perspective. *International Journal of Mobile Communications*, *6*(2), 231-257.

Li, W., & Xu, L. C. (2002). The political economy of privatization and competition: Cross-country evidence from the telecommunications sector. *Journal of Comparative Economics*, *30*(3), 439-462.

Lia, F., & Whalleyb, J. (2002). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy, 26*(2), 451–472.

Lian, Y., Richard, H. W., & Robert, D. (2004). Predicting Customer Behavior in Telecommunications, *Intelligent Systems*, *19*(2), 50-58.

Liang, C., & Dandekar, K. R. (2007). Power Management in MIMO Ad Hoc Networks: A Game-Theoretic Approach. *IEEE Transactions on Wireless Communications, 6*(4), 1164-1170.

Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. *Strategic Management Journal, 9*, 41-58.

Liebowitz, S. J. (2002). *Re-thinking the network economy: the true forces that drive the digital marketplace*. New York: AMACOM.

Liesch, P. W., Welch, L. S., Welch, D., McGaughey, S. L., Petersen, B., & Lamb, P. (2002). Evolving Strands of Research on Firm Internationalization: An Australian-Nordic Perspective. *International Studies of Management & Organization*, *32*(1), 16-35.

Liljander, V., Polsa, P., & Forsberg, K. (2007). Do Mobile CRM Services Appeal to Loyalty Program Customers? *International Journal of E-Business Research*, *3*(2), 24.

Lim B. (1995). Examining the organizational culture and organizational performance link. *Leadership & Organization Development Journal*, *16*(5), 16-21.

Lim, K. Y. (2004). *Communications in Malaysia*. Retrieved March 19, 2005, from http://en.wilipedia.org/wiki/Communications_in_Malaysia.

Lin, H. H., & Wang, Y. S. (2006). An examination of the determinants of customer loyalty in mobile commerce contexts. *Information & Management, 43*(3), 271-282.

Lin, H., Chatterjee, M., Das, S. K., & Basu, K. (2005). ARC: an integrated admission and rate control framework for competitive wireless CDMA data networks using noncooperative games. *IEEE Transactions on Mobile Computing, 4*(3), 243-258.

Lin, H.-C., & Wang, S.-W. (2005). Splitter placement in all-optical WDM networks. *Proceedings of the IEEE GLOBECOM 2005.*

Lin, K.-J. (2008). E-Commerce Technology: Back to a Prominent Future. *Internet Computing*, *12*(1), 60-65. Washington, DC: IEEE Computer Society.

Lin, Y. B, Haung, Y. R., Chen, Y. K., & Chlamtac, I. (2003). Mobility Management: from GPRS to UMTS. *Wireless Comm. and Mobile Computing, 1*(4), 339-359.

Lin, Y. B., & Chlamtac, I. (2000). *Wireless and Mobile Network Architectures.* John Wiley & Sons, Inc.

Lin, Y., Chen, B., & Suter, B. (2007). Robust binary quantizers for distributed detection. *IEEE Transactions on Wireless Communications, 6*(6), 2172-2181.

Ling, R. (2001). *"It is 'in.' It doesn't matter if you need it or not, just that you have it.":* Fashion and the domestication of the mobile telephone among teens in Norway. Oslo: Working Paper, Telenor R&D, Norway.

Ling, R., & Yttri, B. (2002). Hyper-coordination via mobile phone in Norway. In J. E. Katz & M. Aakhus (Eds.), *Perpetual contact.* New York: Cambridge University Press.

Liou, Y. I. (1990). Knowledge acquisition: issues, techniques, and methodology. Paper presented at *Proceedings of the 1990 ACM SIGBDP conference on trends and directions in expert systems,* Orlando, Florida, United States: ACM Press, (pp. 212–36).

Litzkow, M., Livny, M., & Mutka, M. (1998). Condor – A Hunter of Idle Workstations. In *Proceedings of the 8th International Conference of Distributed Computing Systems (ICDCS 1998)* (pp. 104-111). San Jose, CA, USA.

Liu, C. (2005). *Build-in vs. built-on IP address management (IPAM).* White paper, Infoblox. Retrieved June 10, 2006, from http://www.cyberklix.com/PDFs/wp_ipam_05_09_28.pdf

Liu, Q., Safavi-Naini, R., & Sheppard, N. P. (2003). Digital rights management for content distribution. In J. P. Montague & C. Steketee (Eds.), *Conferences in Research and Practice in Information Technology: Australasian Information Security Workshop 2003 (AISW2003))* (Vol. 21). Adelaide: Australian Computer Society.

Loch, C. (2000). Tailoring product development to strategy: case of a European technology manufacturer. *European Management Journal, 18*(3), 246-258.

Lofthouse, H., Yates, M. J., & Stretch, R. (2004). Parlay X Web Services. *BT Technology Journal*, 22(1), 81-86.

Lometti, A. (2006). Transport-MPLS (T-MPLS): An enabling technology for packet transport networks. *Proc. ECOC 2006*, *5*, 91–110.

Longano, S. (2007, May). *T.Net – Trentino in Rete.* Paper presented at the "Bridging the Broadband Gap" conference, Brussels.

Longstaff, F. A., & Schwartz, E. S. (2001). Valuing American Options by Simulation: A Simple Least-Squares Approach. *The Review of Financial Studies, 14*(1), Spring.

Lovelock, C. H., & Yip, G. S. (1996). Developing global strategies for service businesses. *California Management Review, 38*(2), 64-86.

Low, B. (2005). The evolution of China's telecommunication equipment market: A contextual, analytical framework. *Journal of Business and Industrial Marketing, 20*(2), 99-108.

Low, B., & Johnston, W. (2005). Managing ambiguous policies in China's telecommunication market. *Journal of Asia Pacific Business*, *6*(4), 5-30.

Low, B., Johnston, W., & Wang, J. (2007). Securing network legitimacy in China's telecommunication market. *Journal of Business and Industrial Marketing, 22*(2), 97-106.

Lu, G. (2000). Issues and technologies for supporting multimedia communications over the Internet. *Computer Communications*, *23*(14-15), 1323-1335.

Lumpkin, G., & Dess, G. (2004). E-business strategies and internet business models: how the internet adds value. *Organizational Dynamics*, *33*, 161–173.

Luostarinen, R. (1994). *Internationalization of Finnish Firms and their Response to Global Challenges - Research for Action*. Forssa: WIDER Report. The United Nations University/WIDER.

Lyon, T. P., & Huang, H. (1995). Asymmetric regulation and incentives for innovation. *Industrial and Corporate Change, 4*(4), 769-776.

Lyytinen, K., & King, J. L. (2002). Around the cradle of the wireless revolution: the emergence and evolution of cellular telephony. *Telecommunications Policy, 26*(3-4), 97-100.

Ma, W., & Fang, Y. (2002). Two-level forwarding pointer for location management in PCS networks. *IEEE Transactions on Mobile Computing, 1*(1), 32-45.

Maamar, Z. (2003). Commerce, E-Commerce, and M-Commerce: What Comes Next? *Communications of the ACM*, *46*(12), 251-257. New York, NY: ACM Press.

MacDonald, M. (2003). *Peer-to-Peer with VB .NET.* New York: Springer-Verlag.

MacInnes, I. (2005). Dynamic business model framework for emerging technologies. *International Journal of Services Technology and Management, 6*(1), 3-19.

Mackay, M. M., & Weidlich, O. (2007, March 2007). Australian Mobile Phone Lifestyle Index, 3rd Edition, Australian Interactive Media Industry Association (AIMIA), Mobile Content Industry Development Group (MCIDG). Retrieved 8 April, 2007, from http://www.aimia.com.au/mobile

MacKenzie, A., & DaSilva, L. (2006). *Game Theory for Wireless Engineers*. Ft. Collins, CO: Morgan & Claypool.

Maclean, G. (1995). Will fiber optics threaten satellite communications? *Space Policy, 11*, 95-99.

Madhavan, R., & Grover, R. (1998). From embedded knowledge to embodied knowledge: New product development as knowledge management. *Journal of Marketing, 62*(4), 1-12.

Madruga E. L., & Tarouco L. M. R. (1994). Fault management tools for a cooperative and decentralized network operations environment. *IEEE Journal on Selected Areas in Communications*, *12*(6), 1121-1130.

Maesschalck, S., Colle, D., Lievens, I., Picavet, M., & Demeester, P. (2003). Pan-European Optical transport networks: An availability-based comparison. *Photonic Network Communications, 5*, 203-225.

Mahatanankoon, P., & Garcia, J. (2007). Mobile Commerce Adoption Barriers. *Encyclopedia of Mobile Computing and Commerce*. IGI Global.

Maitland, E., & Nicholas, S. (2002). Preface: Modeling Multinationals from Small, Open Economies. *International Studies of Management & Organization*, *32*(1), 3-15.

Majkgård, A., & Sharma, D. D. (1998). Client-Following and Market-Seeking Strategies in the Internationalization of Service Firms. *Journal of Business-to-Business Marketing, 4*(3), 1-41.

Major, M. (2007). Activity-based costing and management: A critical perspective. In Hopper, T., Scapens, R. W., & Northcott, D (Eds.), *Issues in management accounting* (pp. 155-174), 3rd edition. London: Prentice Hall.

Major, M., & Hopper, T. (2005). Managers divided: Resistance and consent to ABC in a Portuguese telecommunications company. *Management Accounting Research*, *16*, 205-229.

Maldoom, D., Marsden, R., Sidak, J. G., & Singer, H. J. (2003). *Competition in broadband provision and its implications for regulatory policy*. Retrieved March 30, 2008, from http://ssrn.com/abstract=463041

Malecki, E. (2002). The economic geography of the Internet's infrastructure. *Economic Geography, 78*, 399-424.

Malhorta, Y. (1998). Deciphering the KM hype, *Learning and Information Management, 21*(4), 58-60.

Malik, O. (2007). 3Q 2007 Broadband Report Card: B-Minus. *GigaOm*, November 13. Retrieved May 8, 2008 from http://gigaom.com/2007/11/13/3q-2007-broadband-report-card-b-minus/.

Malis, A., Pate, P., Cohen, R., Ed., Zelig, D. Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Circuit Emulation over Packet (CEP), IETF, RFC 4842 (2007)

Mallat, N. (2006a). *Consumer and merchant adoption of mobile payments*. Doctoral dissertation. Helsinki School of Economics 2006.

Mallat, N. (2006b). *Exploring Consumer Adoption of Mobile Payments*. Presented at Helsinki Mobility Roundtable, 1-2 June 2006, Helsinki.

Mallenius, S. (2000). *The COPS (Common Open Policy Service) Protocol (Research seminar on IP Quality of Service)*. Helsinki, University of Helsinki, Department of Computer Science.

Mallenius, S. (2000). *The COPS (Common Open Policy Service) Protocol (Research seminar on IP Quality of Service)*. Helsinki, University of Helsinki, Department of Computer Science.

Malmi, T. (1997). Towards explaining activity-based costing failure: Accounting and control in a decentralized organisation. *Management Accounting Research*, *8*, 459-480.

Mammeri, Z., & Lorenz, P. (2004). High-speed networks and multimedia communications, *Lecture Notes in Computer Science (LNCS)*, 3079.

Mangalindan, M., & Rhoads, C. (2007). Sorry, Wrong Number, eBay Says on Skype. *The Wall Street Journal*, October 2, p. B1.

Mannings, R., & Cosier, G. (2001). Wireless everything — unwiring the world. *BT Technol J Vol 19 No 4 October 2001*, *19*(4), 65-76.

Manvi, S. S., & Venkataram, P. (2004). Applications of Agent Technology in Communications: A Review. *Computer Communications*, *27(15)*, 1493-1508.

Manvi, S. S., Kakkasageri, M. S., & Jeremy, P. (2007). Information search and access in vehicular ad hoc networks (VANETs): An agent based approach. In H. R. Arabnia (Ed.), *The International Conference on Communication in Computing (CIC-2007), The 2007 World Congress in Computer Science, Computer Engineering and Applied Computing* (pp. 23-29), CSREA Press: USA.

Manvi, S. S., Kakkasageri, M. S., Jeremy, P., & Alex, R. (2006). Multi agent systems as a platform for VANETs. In H. Nakashima, P. Michael, Wellman, G. Weiss, P. Stone (Eds.), *Autonomous Agents and Multi Agent Systems (AAMAS) (*pp. 35-42), ACM: Japan.

Manyika, R., Roberts, R., & Sprague, K. (2007, December). *McKinsey Quarterly*. Available at: http://yaleglobal.yale.edu/display.article?id=10075. Retrieved, 15 January 2008.

Marcoulides, G. A., & Heck, R. H. (1993). Organizational culture and performance: proposing and testing a model. *Organization Science, 4*(2), 209-25.

Marcus, J. S., & Elixmann, D. (2007). Regulatory Approaches to Next Generation Networks (NGNs): An International Comparison. Unpublished manuscript.

Mark L. (2006). *Change, How to Do it and Make it Work, Tools and Techniques for Managing Change*. A Joint Working Publication.

Mark, G. (2005). Connecting existing phones to VoIP. *Network World, 22*(40), 50.

Mark, H., Catja, P., & Peter, A. (2004). Using Means-End Chains to Build Mobile Brand Communities. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences. Track7, 7*, 70196.3. Washington: IEEE Computer Society.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investment*. New Haven, CN: Yale University Press.

Markus, S., Florian, K., Timo, K., & Florian D. (2005). Secure communication for intersection assistance. In H. Rohling (Ed.), *The Second International Workshop on Intelligent Transportation* (pp. 156-163). IEEE: Germany.

Marra, W., & Schesser, J. (1996). Africa ONE: The Africa optical network. *Communications Magazine*, *34*(2), 50-57.

Martin, L. M., & Matlay, H. (2003). Innovative use of the internet in established small firms: The impact of knowledge management and organizational learning in accessing new opportunities. *Qualitative Market Research: An International Journal, 6*(1), 18-26.

Martini, L. et al., Encapsulation for Transport of ATM, Frame Relay, PPP, HDLC, Ethernet over MPLS, RFC 4717, 4619, 4618, 4448, 4906 (2006)

Marx, L. M. (2006). Economics at the Federal Communications Commission, *Review of Industrial Organization*, *29*(4), 349-368.

Mas, C., Tomkos, I., & Tonguz, O. (2005). Failure location algorithm for transparent optical networks. *IEEE Journal on Selected Areas in Communications, 23*(8), 1508-1519.

Mase, K., & Adjih, C. (2006). *No Overhead auto-configuration OLSR*. IETF draft, February 2006, draft-mase-manet-autoconf-noaolsr-01.txt. Retrieved December 11, 2007, from http://www.tools.ietf.org/html/draft-mase-manet-autoconf-noaolsr-01.

Mase, K., & Weniger, K. (2006). *PDAD-OLSR: Passive duplicate address detection for OLSR*. IETF draft, June 2006, draft-weniger-autoconf-pdad-olsr-01.txt, Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-weniger-autoconf-pdad-olsr-01.

Massey, A. P., Khatri, V., & Ramesh, V. (2005). *From the web to the wireless web: technology readiness and* *usability*. Paper presented at the 38th Hawaii International Conference on System Sciences, Hawaii.

Massey, D. (1993). Power-geometry and a progressive sense of place. In J. Bird, B. Curtis, T. Putnam, G. Robertson, & L. Tickner (Eds.), *Mapping the futures: Local cultures, global change* (pp. 59-69). London: Routledge.

Mathieu, P., & Verrons, M. H. (2002). A generic model for contract negotiation. *Proceedings of the Artificial Intelligence and the Simulation of Behaviour symposium on Intelligent Agents in Virtual Markets (AISB'02), 3*(3), 1-8.

Mathur, S., Sankaranarayanan, L. & Mandayam, N.B. (2006). Coalitional games in receiver cooperation for spectrum sharing. 40th *Annual Conference on Information Sciences and Systems, 1*, 949-954).

Maull, R., Brown, P., & Cliffe, R. (2001). Organizational culture and quality improvement. *International Journal of Operations & Production Management, 21* (3), 302-326.

Maxim, R., & Jean, H. (2005). The security of vehicular ad hoc networks. In Y. W. Law (Ed.), *The Third ACM workshop on Security of ad hoc and sensor networks* (pp. 11-21). ACM: USA.

Maxwell, G., McDougall, M., & Blair, S. (2000). Managing diversity in the hotel sector: the emergence of a service quality opportunity. *Managing Service Quality*, *16*, 367-73.

McBurney, P., Parsons, S., & Green, J. (2002). Forecasting market demand for new telecommunications services: an introduction. *Journal of Telematics and Informatics, 19*(3), 225-249.

McClelland, S. (2006, Tues, May 23. 2006). *WiMAX World Europe: We Want Spectrum, and Good Prices, and, er, LicensesWhat Impediments Are There to WiMAX?* Retrieved, from the World Wide Web: http://www.telecommagazine.com/TechZones/backoffice/

Mccloghrie, K., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., & Waldbusser, S. (1999). Structure of Management Information Version 2 (SMIv2). *Internet Engineering Task Force, Request for Comments: 2578 (STD: 58).*

McCormick, A. (2007). Service-enabled networks aid convergence. *Lightwave, 24*(5).

McDermott & O'Dell (2000). Overcoming the cultural barriers to sharing knowledge. Available at: www.apqc.org/free/articles/km0200/index.htm

McDonagh, D., Bruseberg, A., & Haslam, C. (2002). Visual product evaluation: exploring users' emotional relationships with products. *Applied Ergonomics, 33(1)*, 231-420.

McDonough, C. (1999). The Price of Zoning Revisited: Zoning Issues Raised by the Telecommunications Act of 1996. *Illinois Real Estate Letter*, *13*(1), 1-3.

McDonough, C. (2005) *History of the Wireless Industry in the United States*, report.

McDonough, C. (2007). *The Impact of AWS on the Structure of the Wireless Telephony Industry*. Paper presented at the 2007 Eastern Economics Association Conference, New York, New York.

McDonough, C. (2008). *The Structure of the United States' Wireless Telephone Industry:Theory, Reality and the Role of Government Regulation,* International Telecommunications Society 17th Biennial Conference, Montreal.

McDougall, P., Shane, S., & Oviatt, B. (1994). Explaining the Formation of International New Ventures: The Limits of Theories from International Business Research. *Journal of Business Venturing, 9*, 469-487.

McEvily, S. K., Eisenhardt, K. M., & Prescott, J. E. (2004). The global acquisition, leverage, and protection of technological competencies. *Strategic Management Journal*, *25*. 713–722.

McGinnis, E., & Perkins, D. (1996). *Understanding SNMP MIBs*. New York: Prentice Hall.

McGinnis, E., & Perkins, D. (1996). *Understanding SNMP MIBs*. New Jersey, Prentice Hall.

McGregor, J. (2007, May 14). *Most innovative companies* (p. 60).

McKinion, J. M., Turner, S. B., Willers, J. L., Read, J. J., Jenkins, J. N., & McDade, J. (2004). Wireless technology and satellite internet access for high-speed whole farm connectivity in precision agriculture. *Agricultural Systems*, *81*(2004), 201-212.

McKinion, J. M., Willers, J. L., & Jenkins, J. N. (2007). Wide area wireless network (WAWN) for supporting Precision Agriculture. *Proceedings of the 2007 Beltwide Cotton Production Research Conferences, New Orleans, LA*. Unpaginated CDROM.

McMaster, T., & Wastell, D. (2005). Diffusion-or Delusion? Challenging an IS Research Tradition. . *Information Technology and People, 18*(4), 383–404.

McWalter, W., Thaler, D., & Kessler, A. (2007, December). *RFC 5132 IP Multicast MIB.*

Medhi, D. (2007). Routing management in the PSTN and the Internet: A historical perspective. *Journal of Network and Systems Management, 15*(4), 503-523.

Meier, H., Friedrich, D. R., & Blankensteir, H. (2003). Mobile data: creating sustainable value. *Booz|Allen|Hamilton*, (pp. 1-12).

Meier-Hellstern, K. S., Alonso, E., & O'Neil, D.R. (1992). The use of SS7 and GSM to support high density personal communications. *In Proc. IEEE ICC'92*, Chicago, IL, paper 356.2.

Meister, D. B., & Compeau, D. R. (2002). Infusion of Innovation Adoption: An Individual Perspective. *Proceedings of the ASAC, Winnipeg, Manitoba*.

Melchiors, C., & Tarouco L. M. R. (2000). Troubleshooting network faults using past experience. *Network Operations and Management Symposium*. Honolulu, Hawaii (pp. 549-562).

Melian, J. P., Lopez da Silva, R. A., & Gutiererez, R. A. (2002). *Redes IP de nueva* generation.Communicatciones de Telefonica I+D, No.24

Melody, W. (1999). Telecom reform. *Telecommunications Policy*, *25*, 7 – 34.

Menasce, D., & Almeida, V. (2002). *Capacity planning for Web services: Metrics, Models, and Methods*. Prentice Hall PTR, Upper Saddle River, NJ.

Menasce, D., & Casalicchio, E. (2004). A Framework for Resource Allocation in Grid Computing. In *Proceedings of the 12th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'04)*, (pp. 259-267).

Menelaos, K. P., Fotis, G. C., Iakovos, S. V., & Gennaro, M. (1999). Mobile agent standards and available platforms. *Computer networks, 31(19)*, 1999-2016.

Merchant, A., & Sengupta, B. (1995). Assignment of cells to switches in PCS Networks. *IEEE/ACM Transactions on networking*, *3*(5), 521–526.

Merrett, D. T. (2002). Australia's Emergent Multinationals. *International Studies of Management & Organization*, *32*(1), 109-135.

Meyer-Base, A., & Watzel, R. (1998). Transformation radial basis neural network for relevant feature selection, *Pattern Recognition Letters*, *19*, 1301–1306.

Mezher, T., Abdul-Malak, M. A., & Dayya, M. (1999). Decision support system for managing infrastructure telecommunications projects. *Journal of Management in Engineering*, *15*(1), 74-81.

MIC (2005). *Interconnection Criteria of Telecommunications Facilities.* Public Announcement No. 2005-5 (in Korean).

Michaela, O. (2002). An Adaptive Negotiation Model for Agent Based Electronics Commerce. S*tudies in Informatics and Control, 11(3)*, 45-46.

Mier, E. E. (2007). SIP: A mid-course assessment. *Business Communications Review*, *37*(3), 28.

Miles, M. B., & Huberman, A. M. (1994). *An Expanded Sourcebook: Qualitative Data Analysis*. Thousands Oaks: Sage.

Milgrom, P. (2004). *Putting auction theory to work*. Cambridge: Cambridge University Press.

Miller, D. R. (2000). Rapid ethnography: time deepening strategies for HCI field research. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques* (280-286). New York, NY: ACM.

Miller, J. G., & Vollmann, T. E. (1985). The hidden factory. *Harvard Business Review*, September-October, 142-150.

Mills, J., & Platts, K. (2003). Applying Resource Based Theory, Methods, Outcomes and Utility for Managers, *International Journal of Operations and Production Management*, *23*(2), 148-166.

Minagar, A., & Premaratne, M. (2006). Cost optimal allocation of amplifiers and DCMs in WDM ring networks. *Optic Express, 14*(22), 10278-10291.

Minar, N., & Hedlund, M. (2001). A Network of Peers: Peer-to-Peer Models Through the History of the Internet. In A. Oram (Ed.), *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology* (pp. 3-20) Sebastopol, CA, USA. O'Reilly.

Minch, R. P. (2004). *Privacy issues in location-aware mobile devices.* Paper presented at the Proceedings of the 37th Hawaii International Conference on System Sciences, January 5-8

Mingay S., & Bittinger S. (2002). Combine CobiT and ITIL for Powerful IT Governance. In *Research Note*, TG-16-1849, Gartner.

Minges, M. (2005). Is the Internet Mobile? Measurements from the Asia-Pacific region. *Telecommunications Policy 29 (2005), 113–125.*

Ministry of Construction & Transportation (2006). from http://www.moct.go.kr.

Ministry of Information and Communication (2006). from http://www.mic.go.kr.

Ministry of Information and Communication (2005). *Plan for Intelligent Home Network Service Distribution and Industry Nurturing Plan*, Government Report.

Mir, R. M. (2008). http://www.cs.wustl.edu/~jain/cis788-97/ftp/satellite_data.pdf. (Verified January 23, 2008)

Mishra, A. R. (2004). Fundamentals of Cellular Network Planning and Optimisation. John Wiley & Sons.

Mishra, S. M., Sahai, A., & Brodersen, R. W. (2006). Cooperative sensing among cognitive radios. *IEEE International Conference on Communications, 1*, 1658-1663.

Mislove, A., & Druschel, P. (2004). Providing administrative control and autonomy in structured peer-to-peer overlays. In *Proceedings of the 3rd International Workshop on Peer-to-Peer Systems (IPTPS04)*. San Diego, CA.

Mitola, J., III, & Maguire, G. Q., Jr. (1999). Cognitive radio: Making software radios more personal. *IEEE Personal Commununications*, *6*(4), 13-18.

Miyaoku, K., Higashino, S., & Tonomura, Y. (2004). C-Blink: A Hue-Difference-Based Light Signal Marker for Large Screen Interaction via any Mobile Terminal. *17th Annual ACM Symposium on User Interface Software and Technology*. New York, NY: ACM Press.

Moerdijk, A., & Klostermann, L. (2003). Opening the networks with OSA/Parlay APIs: Standards and aspects behind the APIs. *IEEE Network Magazine*, 17(3), 58-64.

Moffett, S., McAdam, R., & Partkinson, S. (2003). An empirical analysis of knowledge management applications. *Journal of Knowledge Management*, *7*(3), 6-26.

Mohsin, M., & Prakash, R. (2002, October). IP address assignment in a mobile ad-hoc network, in Proc. *IEEE Military Communications Conference (MILCOM 2002)*, *2*, 856-861.

Mole, K. F., Ghobadian, A., O'Reagan, N., & Liu, J. (2004). The Use and Deployment of Soft Process Technologies within UK Manufacturing SMEs: An empirical Assessment Using Logit Models. *Journal of Small Business Management*, *42*(3), 303-324.

Molina, A., Athanasiadou, G. E., & Nix, A. R. (2000). The automatic location of base-stations for optimized cellular coverage: A new combinatorial approach. *IEEE Vehicular Technology Conference* (pp. 606-610).

Monge, P., & Contractor, N. (2003). *Theories of Communication Networks*. New York, NY: Oxford Press.

Monge, P., & Eisenberg, E. (1987). Emergent communication networks. In F. Jablin, L. Putnam, K. Roberts, & C. O'Reilly (Eds.), *Handbook of Organizational Communication* (pp. 304-342). Norwood, NJ: Ablex.

Montes, F. J. , Moreno, A. R., & Fernandez, L. M. (2004). Assessing the organizational climate and contractual relationship for perceptions of support for innovation. *International Journal of Manpower, 25*(2), 167-180.

Montgomery, W. (1983). Techniques for Packet Voice Synchronization. *IEEE Journal on Selected Areas in Communications, 1*(6), 1022–1028.

Moon, S. B., Kurose, J., & Towsley, D. (1998). Packet audio *playout* delay adjustment: performance bounds and algorithms. *Multimedia Systems, 6*, 17–28.

Morgan, G. (1989). *Creative Organization Theory: A Resource Handbook*. London: Sage.

Morgan, G. (1997). *Images of Organisation*. London: Sage Publications.

Moritz, S. (2007). AT&T, EchoStar Deal Nears. *TheStreet.com*, November 19.

Morkel, A., & Osegowitsch, T. (1999). The international competitiveness of Australian business. In G. Lewis, A. Morkel, G. Hubbard, S. Davenport & G. Stockport (Eds.), *Australian and New Zealand Strategic Management: Concepts, Context and Cases*. Sydney: Prentice Hall Australia Pty Ltd.

Morlot, F., Fourestié B., & Elayoubi, S. E. (2007). Optimizing the date of an upgrading investment in a data network. *Proceedings of the IEEE Vehicular Technology Conference IEEE VTC-fall 2007*, Baltimore, October 2007.

Morris, A. C. (2005). Spectrum auctions: Distortionary input tax or efficient revenue instrument? *Telecommunications Policy, 29*(9-10), 687-709.

Morrissey, P. (2005). Skype: Unsafe in any enterprise. *Network Computing, 16*(23), 22.

Motion Picture Association. Belgian & Swiss Authorities Break Razorback2 [press release]. MPA, 21 February 2006.

Mozer, M. C., Wolniewicz, R., & Grimes, D. B. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks, 11*(3), 690–696.

Mueller, M. L. (1993). *Telephone companies in paradise: A case study in telecommunications deregulation*. New Brunswick, NJ: Transaction Publishers.

Multiprotocol Label Switching, IETF Working group <www.ietf.org/html.charters/mpls-charter.html>

Mun, J. (2002). *Real Options Analysis: Tools and techniques for valuing strategic investments and decisions*. Finance: Wiley.

MuniWireless. (2008). Retrieved January 30, 2008, from www.muniwireless.com

Murat, C., Martin, M., Bernd, R., & Andreas, L. (2005). Collection of dedicated information in vehicular ad hoc networks. In L. Katz & R.M. Sorrentino (Eds.), *12th World Congress on Intelligent Transport Systems* (pp. 195-207). Cognizant: U.S.A.

Mustonen-Ollila, E., & Lyytinen, K. (2003). Why Organizations Adopt Information System Process Innovations: A longitudinal study using diffusion of innovation theory,. *Information Systems Journal, 13*(3), 275-292.

Muzzi, C., & Kautz, K. (2004). Information and communication technologies diffusion in industrial districts. In J. Damsgaard & H. Z. Henriksen (Eds.), *Networked Information Technologies: Diffusion and Adoption* (pp. 19-38). Dordrecht: Kluwer Academic Publishers.

Mylonopoulos, N. A., & Sideris, I. A. (2006). Growth of value added mobile services under different scenarios of industry evolution. *Electronic Markets, 16*(1), 28-40.

Nabrzysbi, J., Schopf, J., & Weglarz, J. (2004). *Grid Resource Management*. Boston, MA: Kluwer Academic Publication.

Narbutt, M., & Murphy, L. (2004). Improving voice over IP subjective call quality. *IEEE Communications Letters, 8*(5), 308-310.

Nardi, B. (1997). The use of ethnographic methods in design and evaluation. In M. Helander, T. K. Landauer, and P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed, pp. 361-366). Amsterdam, Netherlands: Elsevier.

Nash, D. A., & Ragsdale D. J. (2001). Simulation of self-similarity in network utilization patterns as a precursor to automated testing of intrusion detection systems. *IEEE Transactions on Systems, Man and Cybernetics, 31,* I(4), 327–331. Dept. of Electr. Eng. & Comput. Sci., US Mil. Acad., West Point, NY, USA.

Nash, J. (1951). Non-cooperative Games. *Annals of Mathematics, 5*(2), 286-295.

Nath, B., Reynolds, F., & Want, R. (2006). RFID Technology and Applications. *Pervasive Computing*, *5*(1), 22-24. Washington, DC: IEEE Computer Society.

National Information Society Agency (2003). *Information Technology Advancement, Planning Report on Next-generation IT Growth Engine (Home Network Part)*, IITA.

NCTA (National Cable & Telecommunications Association). (2007). *Statistic*s. Washington, DC: NCTA. Retrieved January 15, 2008, from http://ncta.com/ContentView.aspx?contendID=54.

Neisse, R., Granville, L. Z., Almeida, M. J., & Tarouco, L. (2003). A Dynamic SNMP to XML Proxy Solution. In *8th IFIP/IEEE International Symposium on Integrated Network Management* (pp 481-484).

Neisse, R., Vianna, R. L., Granville, L. Z., Almeida, M. J. B., & Tarouco, L. M. R. (2004). Implementation and Bandwidth Consumption Evaluation of SNMP to Web Sevices Gateways. In *IEEE/IFIP Network Operations and Management Symposium* (pp. 715-728).

Nekovee, M. (2006). Dynamic spectrum access – Concepts and future architectures. *BT Technology Journal*, *24*(2), 111-116.

Nemcek, A. (2002). *Motorola receives US$446 millions CDMA 1X contracts from China Unicom*. Motorola Press Release.

Nesbitt, K. (2002). *Designing a knowledge management system*. Retrieved March 19, 2005, from http://academic.edu:2081/products/faulknerlibrary/00018382.htm.

Neuman, W. L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches* (4th ed.). Boston: Allyn and Bacon.

New York Times (1999). *Holy War Over the Future of Wireless.*

Newman, A. L., & Bach, D. (2004). Privacy and regulation in a digital age. In B. Preissl, H. Bouwman & C. Steinfield (Eds.), *E-Life After the Dot.Com Bust*. Heidelberg: Physica-Verlag.

Newman, K. L., & Nollen, S. D. (1996). Culture and congruence: The fit between management practices and national culture. *Journal of International Business*, *27*, 753-79.

Ng, G., & Ong, K. (2000). Using a qualitative probabilistic network to explain diagnostic reasoning in an expert system for chest pain diagnosis. *Computers in Cardiology* (pp. 569-572).

Ng, K., & Liu, H. (2001). Customer retention via data mining, *Issues on the Application of Data Mining. Artificial Intelligence Review*, *14*(6), 569–590.

Ng-Kruelle, G., Swatman, P. A., Rebne, D. S., & Hempe, J. F. (2001). The price of convenience: Privacy and mobile commerce. *Quarterly Journal of Electronic Commerce*, *3*(4), 273-285.

Ngo, M. H., & Krishnamurthy, V. (2007). Game theoretic cross-layer transmission policies in multipacket reception wireless networks. *IEEE Transactions Signal Processing, 55*(5-1), 1911-1926.

Nichols, K., & Carpenter, B. (2001). *Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification*, IETF, RFC 3086.

Nickerson, M. (2005). History Calls: Delivering Automated Audio Tours to Visitors' Cell Phones. *2005 IEEE Conference on Information Technology: Coding and Computing.* Washington, DC: IEEE Computer Society.

Nie, N., & Comaniciu, C. (2006). Adaptive channel allocation spectrum etiquette for cognitive radio networks. *Springer Mobile Networks and Applications*, *11*(6), 779-797.

Nie, N., & Comaniciu, C. (2006). Adaptive channel allocation spectrum etiquette for cognitive radio networks. *Springer Mobile Networks and Applications*, *11*(6), 779-797.

Nielsen, E. H., & Rao, M. V. H. (1987). The strategy-legitimacy nexus: A thick description'. *Academy of Management Review*, *12*, 523–33.

Nielsen, J. (2000). *Designing Web Usability.* Indianapolis, IN: New Riders.

Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods.* New York, NY: John Wiley & Sons.

Nigay, L. P., Salembier, P., Marchand, T., Renevier, P., & Pasqualetti, L. (2002). Mobile and Collaborative Augmented Reality: A Scenario Based Design Approach. In *Proceedings of Mobile HCI 2002* (pp. 241-255).

Nielsen Media Research. (2006). *Cable TV audience: 2006 annual report.* Retrieved on January 22, 2008 from http://www.journalism.org/node/509

Nixon, J., Sharples, S., & Jackson, M. (2007). Presenting spatial information on a mobile device: Differences in workload and performance. In *Proceedings of Human Factors and Ergonomics Society (HFES), 51st Annual Meeting.* Santa Monica, SA: Human Factors and Ergonomics Society.

Niyato et al. (2007, February). IEEE 802.16/WiMAX-based Broadband Wireless Access and its Application for Telemedicine/e/Health services. *IEEE Wireless Communications*, pp. 72-83.

Niyato, D., & Hossain, E. (2007). A game-theoretic approach to competitive Spectrum sharing in cognitive radio networks. *IEEE Wireless Communications and Networking Conference, 1*, 16-20.

Niyato, D., & Hossain, E. (2007). Integration of IEEE 802.11 WLANs with IEEE 802.16-based multihop infrastructure mesh/relay networks: A game-theoretic approach to radio resource management. *IEEE Network*, *21*(3), 6-14.

Noam, E. (2002). *How telecom is becoming a cyclical industry, and what to do about it.* Presentation at Columbia Institute for Tele-Information, Conference on Telecommunications Volatility, April 2002, Columbia University, New York.

Noam, E. (1999). *Telecommunications in Africa*. New York: Oxford.

Noam, E. (2003). The effect of deregulation on market concentration: an analysis of the Telecom Act of 1996 and the industry meltdown. *The Columbia Science and Technology Law Review*, *4*, 1-12.

Noam, E. (2003b). Deregulation and market concentration: An analysis of post-1996 consolidations. *Federal Communications Law Journal*, *58*, 539-550.

Noam, E. (2004). The Internet: Still Wide Open and Competitive? Oxford Internet Institute, Internet Issue Brief No. 1, August 2003.

Noam, E. M. (2001). *Interconnecting the network of networks.* Cambridge, MA: MIT Press.

Nohria, N. (1992). *Is a Network Perspective a Useful Way of Studying Organizations?* In Nohria and Eccles (Eds.), *Networks and Organizations.* Boston, MA: Harvard Business School Press.

Nokia Siemens Parlay @vantage, at http://www.nokiasiemensnetworks.com

Nokia Telecommunications. (1999). *The Demand for Mobile Value Added Services: A Market Study*, available at http://www.nokia.com/press/background/pdf/study-vas.pdf, 1999.

Nolan, P. (2001). *China and the global economy.* Palgrave Publication.

Nonaka, I. (1998). The knowledge-creating company. *Harvard Business Review on Knowledge Management*. Boston, MA: Harvard Business School Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Noreen, E. (1991). Conditions under which activity-based cost systems provide relevant costs. *Journal of Management Accounting Research*, *3*, Fall, 159-168.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York, NY: Basic Books.

North, P. (1990.) *Institutions, institutional change and economic performance*. Cambridge, UK: Cambridge University Press:

Nuechterlein, J., & Weiser, P. (2005). *Digital Crossroads*. Cambridge, MA: The MIT Press.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005). Explaining intention to use mobile chat services: Moderating effects of gender. *Journal of Consumer Marketing, 22*(5), 247-256.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005). Intentions to Use Mobile Services: Antecedents and Cross-Service Comparisons. *Journal of the Academy of Marketing Science*, *33*(3), 1-17.

O'Dell, C., & Grayson, J. (1999). *If only we know what we know: The transfer of internal knowledge and best practice*. New York: Simon and Schuster.

O'Hara-Devereaux, M., & Johansen, R. (1994). *Globalwork*. San Francisco, CA: Jossey-Bass.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *Journal of Operational Research, 32*, 393-404.

O'Regan, N., & Ghobadian A. (2004). Short- and long-term performance in manufacturing SMEs. Different targets, different drivers. *International Journal of Productivity and Performance Management*, *53(5)*, 405-424.

O'Reilly & Chatman, (1996). Culture as Social Control: Corporations, Cults, and Commitment, *Research in Organizational Behavior, 18*, pp. 157 – 200

O'Reilly, T. (2005). *What Is Web 2.0. Design patterns and business models for the next generation of software*. Retrieved March 31, 2007, from http://www.oreillynet.com/ pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

OASIS. (2006). Reference Model for Service Oriented Architecture 1.0. *Committee Specification 1*, Available from http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf

Odlyzko, A., Internet traffic growth: Sources and implications, Proc. SPIE Vol. 5247 (2003) 1-15

OECD. (2007). *Glossary of Statistical Terms*. Retrieved October 13, 2007, from http://stats.oecd.org/glossary/search.asp

OECD (2006). *Internet Traffic Exchange: Market Developments and Measurement of Growth*. DSTI/ICCP/TISP(2005)11/Final, 05 April, retrieved Mar. 11, 2006, from http://www.oecdorg/dataoecd/25/54/36462170.pdf

OECD (2007). The Influence of Market Developments and Policies on Telecommunication Investment. *DSTI/ICCP/CISP 6*.

OECD (2004). Performance Indicators for Public Telecommunication Operators. Paris. Original OECD report available online at: http://www.oecd.org/dsti/sti/it/cm/prod/ICCP22.htm

OECD (2006), *VoIP: Developments in the market*. Working Party on Telecommunication and Information Services Policies, OECD, Paris.

OECD Economics Department Working Papers (2001). OECD Economics Department in its series by Olivier Boylaud and Giuseppe Nicoletti. Regulation, market structure and performance in telecommunications, 2001.

OFCOM (2006). Regulatory challenges posed by next generation access networks. Public discussion document.

OFCOM (2006). Review of the wholesale broadband access markets 2006/07. Public consultation document.

OFCOM (2007). Future broadband policy approach to next generation access. Public consultation document.

Office of the Telecommunications Authority (OFTA) of the Hong Kong Special Administrative Region, Statistics of Internet Traffic Volume (Customer Access via Broadband Networks) (2008) <www.ofta.gov.hk/en/index.html>

OFTEL (2001). Effective Competition Review of Internet Connectivity.

Oh, S., Ahn, J., & B. Kim (2003). Adoption of broadband Internet in Korea: the role of experience in building attitudes. *Journal of Information Technology, 18*, 267-280.

Oh, S., Yang, S., Kurnia, S., Lee, H., Mackay, M. M., & O'Doherty, K. (2008). The characteristics of mobile data users in Australia. *International Journal of Mobile Communications, 6*(2), 217-230.

Oh, Y. J., Ju, H. T., Choi, M. J., & Hong, J. W. (2002). Interaction Translation Methods for XML/SNMP Gateway. In *13h IFIP/IEEE International Workshop on Distributed Systems: Operations and Management* (pp. 54-65).

Ohmae, K. (1989). Managing in a Borderless World. *Harvard Business Review, 67*(3), 152-161.

Ohmori, S., Yamao, Y., & Nakajima, N. (2000). The future generations of mobile communications based on broadband access technologies. *IEEE Communications Magazine, 38*(12), 134- 142.

Ohrtman, F., & Roeder, K. (2003). *Wi-Fi Handbook: Bulding 802.11b Wireless Networks.* New York, NY: McGrawHill.

Oie, Y., Sueda, T., Murata, M., & Miyahara, H. (1991). Survey of switching techniques in high-speed networks and their performance. *International Journal of Satellite Communications, 9*, 285-303.

Okuno, S., & Okuda, M. (2003). Trends toward Globalization of the TL 9000 Quality Management System for the Telecommunications Industry. *NTT Technical Review, 1*(6), 92-95.

Oldale, A., & Padilla, A. J. (2004). *From State Monopoly to the "Investment Ladder": Competition Policy and the NRF.* In *Swedish Competition Authority, The Pros and Cons of Antitrust in Deregulated Markets*, 51-77.

Olley, G., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*, 1263-1297.

Ono, R., & Aoki, K. (1998). Convergence and new regulatory frameworks. *Telecommunication Policy, 22*(11), 817-838.

OPAALS. (2008). *Open Philosophies for Associative Autopoietic digitaL ecosystemS*. Retrieved January 30, 2008, from http://www.opaals.org

OpenGroup. (2008). Service Oriented Architecture (SOA). Available from http://www.opengroup.org/projects/soa/

Openreach (2007a). Openreach services delivered over Fibre To The Premises (FTTP). Consultation document.

Openreach (2007b). Product definition overview: Generic Ethernet Access delivered over Fibre to the Premises.

Oppenheim, J. (1991). In B. G. Cole (Ed.), *After the breakup: Assessing the new post-AT&T divestiture era*, (pp. 310-315). New York: Columbia University Press.

Orfali, R., & Harkey, D. (1998). *Client/Server Programming with Java and CORBA.* John Wiley and Sons, New York, USA, 2nd edition.

Orihata, M., & Watanabe, C. (2000). Evolutional dynamics of product innovation: the case of consumer electronics. *Technovation, 20*(8), 437-449.

Orlikowski, W. J., & Gash, D. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems, 12*(2), 174-207.

Osborne, M. J. (2004). *An introduction to Game Theory.* New York, NY: Oxford University Press.

OSI (1987). Specification of Abstract Syntax Notation One (ASN.1). *International Standard* 8824.

Osterwalder, A., & Pigneur, Y. (2002). Business models and their elements. Paper presented at *the International Workshop on Business Models*, Lausanne, Switzerland.

Osterwalder, A., Ondrus, J., & Pigneur, Y. (2005). *Skype's Disruptive Potential in the Telecom Market: A Systematic Comparison of Business Models.* Retrieved 07.04., 2008, from http://www.hec.unil.ch/yp/Pub/05-skype.pdf

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. *Proceedings of CHI 2005.* New York, NY: ACM press.

Overdorf, M., & Christensen, C. M. (2000). Meeting the Challenge of Disruptive Technology. *Harvard Business Review*, 67-76.

Ovum (2006). Virtual Mobile Services: Strategies for Fixed and Mobile Operators. Ovum Ltd. 2006, by John Matthews, Susan Sweet. http://www.ovum.com.

Owen, B. (2001). *The Internet challenge to television*. Cambridge, MA: Harvard University Press.

Owen, B., & Rosston, G. (2003). Local Broadband Access: *Primum Non Nocere* or *Primum Processi?* A Property Rights Approach. *Joint Center AEI-Brookings Joint Center for Regulatory Studies, 03*-19, p.7.

Pagani, M. (2003). *Multimedia and interactive digital TV: Managing the opportunities created by digital convergence*. Hershey, PA: Idea Group Publishing.

Pailer, R., Stadler, J., & Miladinovic, I. (2003). Using PARLAY APIs Over a SIP System in a Distributed Service Platform for Carrier Grade Multimedia Services. *Wireless Networks,* 9(4), 353-363.

Pandurang, K., Arati, B., & Wade, T. (2006). An identity-based security framework For VANETs. In W. Holfelder, D. Chrysler, & D. B. Johnson, (Eds.), *The Third international workshop on Vehicular ad hoc networks* (pp. 94-95), ACM: USA.

Pang, B., Zhang, D., Li., N., & Kuanquan, W. (2004). Computerized tongue diagnosis based on Bayesian networks. *IEEE Transactions on Biomedical Engineering, 51*(10), 1803-1810.

Papadakis, A. E., Chaniotakis, E .S. ,Giannakakis, P. E., Tselikas, N. D., & Venieris, I. (2004). Parlay-based service provision in circuit- and packet-switched telecommunications networks. *International Journal of Communication Systems. 17*(1), 63-83.

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on human-computer interaction with mobile devices and services (*pp. 203-210). New York, NY: ACM press.

Park H., Ribiere V. & Schulte W. D. Jr., (2004). Critical attributes of organizational culture that promote knowledge management technology implementation success. *Journal of Knowledge Management, 8*(3), 106-117

Park, S. H., & Russo, M. V. (1996). When companies eclipse cooperation: an event history analysis of joint venture failure. *Management Science, 42*, 875-90.

Parsons, P. (2003). The evolution of the cable-satellite distribution system. *Journal of Broadcasting & Electronic Media, 47*(1), 1-17.

Parsons, P., & Frieden, R. (1998). *The cable and satellite television industries.* Needham Heights, MA: Allyn & Bacon.

Pascal R. T., & Athos A. G. (1982). *The Art of Japanese Management.* Harmondsworth: Penguin Books.

Pascoe, J., Ryan, N. S., & Morse, D. R. (2000). Using While Moving: HCI Issues in Fieldwork Environments. *ACM Transactions on Computer Human Interaction*, *7*(3). 417-437.

Patterson, B. (August 16, 2007). Three ways to get HDTV programming. *CNET HDTV World*. Retrieved January 24, 2008, from http://www.cnet.com/4520-7874_1-5108854-2.html

Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.). London: Sage Publications.

Pavitt, K. (1991). Key characteristics of the large innovating firm, *British Journal of Management, 2*, 41-50.

Pavlik, J. (1998). *New media technology: Cultural and commercial perspectives*. Needham Heights, MA: Allyn & Bacon.

Pavlou, G., Flegkas, P., Gouveris, S., & Liotta, A. (2004). On Management Technologies and the Potential of Web Services. *IEEE Communications*, *42*(7).

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly*, *30*(1), 115–143.

Pavlou, P. A., & Gefen, D. (2005). Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating role. *Information Systems Research*, *16*(4), 372–399.

Payne, S. J. (2003). Users' mental models: The Very Ideas. In J. M. Carroll (Ed.), *HCI Models, Theories and Frameworks: Toward a multidisciplinary science*. San Francisco, CA: Morgan Kaufman.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, California: Morgan Kaufmann.

Pedersen, P. E. (2001). An adoption framework for mobile commerce. *Proceedings of the 1st. IFIP Conference of E-Commerce*, Zürich, Switzerland, October 3-5, 2001.

Pedersen, P. E. (2005). Adoption of mobile Internet services: An exploratory study of mobile commerce early adopters. *Journal of Organizational Computing and Electronic Commerce*, *15*(3), 203-221

Pedersen, P. E., & Thorbjørnsen, H. (2003). Adoption of Mobile Services. *Model development and Cross-Service study*. Norwegian School of Economics and Business Administration.

Pedersen, P. E., Methlie, L. B., & Thorbjørnsen, H. (2001). *Understanding mobile commerce end-user adoption: A triangulation perspective and suggestions for an exploratory service evaluation framework* . Presented at HICSS-35, Hawaii, US, Jan 7-10, 2002

Peha, J., Lehr, W., & Wilkie, S. (2007). The State of the Debate on Network Neutrality. *International Journal of Communications*, *1*(1), 709-716.

Pekar, P., & Allio, R. (1994). Making alliances work. Guidelines for success. *Long Range Planning, 27*(4), 54-65.

Peng, M., & Wang, W. (2007). A Unified Architecture and Key Techniques for Interworking between WiMAX and Beyond 3G/4G Systems. *Wireless Personal Communications.*

Pennings, J. M., & Harianto, F. (1992). The diffusion of technological innovations in the commercial banking industry. *Strategic Management Journal, 13*, 29-46.

Peppard, J., & Rylander, A. (2006). From Value Chain to Value Netork: Insight for Mobile Operators. *European Management Journal*, *24*(2-3), 128-141.

Peristeras, V., & Tarabanis, K. (2004, 25-27 October). *Governance Enterprise Architecture (GEA): Domain Models for e-Governance*. Paper presented at the 6th ACM International Conference on Electronic Commerce (ICEC'04), Delft, The Netherlands.

Perkins, C., Malinen, J., Wakikawa, R., Belding-Royer, E., & Sun, Y. (2001). *IP address auto-configuration for Ad Hoc Networks.* IETF draft, November 2001, draft-perkins-manet-autoconf-01.txt. Retrieved December 14, 2007, from http://tools.ietf.org/id/draft-perkins-manet-autoconf-01.txt.

Perrow, C. (1961). Organizational prestige. *American Journal of Sociology, 66*, 335–341.

Peteraf, M. A. (1993). The cornerstone of competitive advantage: a resource based view. *Strategic Management Journal*, *3*(3), 179-191.

Petkovic Y., & De Coster, R. (2000). *Broadband Access: New Business Models.* London: Ovum Ltd.

Pew, R. W., & Mavor, A. S. (2007). *Human-System Integration in the System*

Pfeffer, J., & Salancik. (1978). *The External Control of Organisations: A Resource Dependance Perspective*. New York: Harper & Row.

Pfeffer, J., & Veiga, J. F. (1999). Putting people first for organizational success. *Academy of Management Executive, 13*(2), 37-48.

Pfeifenberger, J., & Houthakker, H. (1998). Competition to international satellite communications services. *Information Economics and Policy*, *10*, 403-430.

Picard, R. (2002). *The economics and financing of media companies*. Bronx, NY: Fordham University Press.

Picker, R. (2005). SBC and Verizon: Judgment Day. *The University of Chicago Law School Faculty Blog*, November 3. Retrieved November 3, 2005 from http://uchicagolaw.typepad.com/faculty/2005/11/sbc_and_verizon.html.

Pierre, S., & Houeto, F. (2002). A tabu search approach for assigning cells to switches in cellular mobile networks. *Computer Communications, 25*, 464-477.

Pikkarainen, H. (2001). 3G market profile: existing services for 3G. *Nokia Networks*, (pp. 1-19).

Pillalamarri, S., & Ghosh, S. (2005). High-speed networks: definition and fundamental attributes. *Computer Communications*, *28*(8), 956–966.

Pilotta, J., Widman, T., & Jasko, S. (1988). Meaning and action in the organizational setting: An interpretive approach. *Communication Yearbook*, *11*, 310-334.

Pindyck, R. (2007). Mandatory unbundling and irreversible investment in telecom networks. *Review of Network Economics*, *6*(3), 274-298.

Pindyck, R. (2007). Mandatory Unbundling and Irreversible Investment in Telecom Networks. *Review of Network Economics*, *6*(3), 274-298.

Pioro, M., & Medhi, D. (2004). *Routing, Flow, and Capacity Design in Communication and Computer Networks*. The Morgan Kaufmann Series in Networking.

Pitkänen, O. (2006). *Legal and regulation framework specification: competence within mobile families and ad-hoc communities (IST-2004-511607 MobiLife)*. Finland: Information Society Technologies.

Pitkow, J., Schutze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., et al. (2002). Personalized search. *Communications of the ACM, 45*(9), 50-55.

Plot, C. R., & Salmon, T. C. (2004). The simultaneous ascending auction: Dynamics of price adjustment in experiments and in the UK 3G spectrum auction. *Journal of Economic Behaviour and Organization*, *53*(3), 353-383.

Podnar, H., & Skorin-Kapov, J. (2002a). An application of a genetic algorithm for throughput optimization in non-broadcast WDM optical networks with regular topologies. *Mathematical Communications, 7,* 45-49.

Podnar, H., & Skorin-Kapov, J. (2002b). A genetic algorithm for optimizing throughput in non-broadcast WDM optical networks. In A. Abraham, & M. Koeppen (Eds.), *Hybrid Information Systems* (pp. 323-338). Heidelberg: Physica Verlag.

Podnar, H., Skorin-Kapov, J., & Skorin-Kapov, D. (2002). Network cost minimization using threshold-based discounting. *European Journal of Operational Research*, *137*, 371-386.

Pollet, J. (2002). Developing a solid SCADA security strategy. *2nd. ISA/IEEE Sensors for Industry Conference*, 2002, (pp. 148-156). PlantData Technologies.

Pool, S. W. (2000). The learning organization: motivating employees by integrating TQM philosophy in a supportive organizational culture. *Leadership & Organisation Development Journal, 21(*8), 373-8.

Poole, I. S. (1990). *Technologies without boundaries*. Cambridge, MA: Harvard U. Press.

Poole, P. P. (1998). Words and deeds of organizational change. *Journal of Managerial Issues, 10*(1), 45-59.

Popovski, P., Yomo, H., Nishimori, K., Di Taranto, R., & Prasad, R. (2007). Opportunistic interference cancellation in cognitive radio systems. *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 472-475.

Porter, M. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York, NY: Free Press.

Porter, M. E. (1985). *Competitive advantage: creating and sustaining superior performance*. New York: Free Press.

Postel, J. Standard for the transmission of IP datagrams over experimental Ethernet (IEEE 802) networks, IETF, RFC 895, 1042 (2005)

Prabhakar, G., Rastogi, R., & Thottan, M. (2005). OSS architecture and requirements for VoIP networks. *Bell System Technical Journal, 10*(1), 31-45.

Pras, A., Schowalder, J., Burgess, M., Festor, O., Pérez, G. M., Stadler, R., & Stiller, B. (2007). Key research challenges in network management. *IEEE Communications Magazine, 45*(10), 104-110.

Pravda, I., & Vodrazka, J. (2007). Voice quality planning for NGN including mobile networks. *IFIP International Federation for Information Processing, 245*, 376-383.

Presuhn, R., Case, J., Mccloghrie, K., Rose, M., & Waldbusser, S. (2002). Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP). *Internet Engineering Task Force. Request for Comments: 3416 (STD: 62).*

Price, J., & Javidi, T. (2007). Leveraging downlink for efficient uplink allocation in a single-hop wireless network. *IEEE Transactions on Information Theory, 53*(11), 4330-4339.

Prieger, J. E. (2001). Telecommunications regulation and new services: A case study at the state level. *Journal of Regulatory Economics, 20*(3), 285-305.

Prieger, J. E. (2002). A model for regulated product innovation and introduction with application to telecommunications. *Applied Economics Letters, 9*(10), 625-629.

Prieger, J. E. (2003). The supply side of the digital divide: Is there equal availability in the broadband Internet access market? *Economic Inquiry, 41*(2), 346–363.

Prieger, J. E. (2004). Regulation, innovation, and the introduction of new telecommunications services. *Review of Economics and Statistics, 84*(4), 704-715.

Prieger, J. E. (2007). Regulatory delay and the timing of product innovation. *The International Journal of Industrial Organization, 25*(2), 219-236.

Prieger, J. E. (2008). Product innovation, signaling, and endogenous regulatory delay. *Journal of Regulatory Economics, 34*(2), in press.

Prieger, J. E., & Heil, D. (2008). Is regulation a roadblock on the information highway? In I. Lee (Ed.), *Handbook of Research on Telecommunications Planning and Management for Business*. Hershey, PA: IGI Global (pp. X-Y).

Prieger, J. E., & Lee, S. (2008). Regulation and the deployment of broadband. In Y.K. Dwivedi, et al. (Eds.), *Handbook of Research on Global Diffusion of Broadband Data Transmission* (pp. 241-259). Hershey, PA: IGI Global.

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences*. San Francisco: W.H. Freeman & Co.

Probe Research (2003). *Global Internet Traffic-2003.*

Proff, H. (2002). Business unit strategies between regionalisation and globalisation. *International Business Review, 11*(2), 231-250.

Proteus: IP address management built right. (2007). White paper. BlueCat Networks. Retrieved September 14, 2007, from http://download.101com.com/techlibrary/bluecat/bluecat_proteus_ip_address_mgmt.pdf

Protter, P., Clément, E., & Lamberton, D. (2002). An analysis of a least squares regression method for American option pricing. *Finance and Stochastics, Springer, 6*(4), 449-471.

Psimmenos, I. (2007). The EC single market and employees' participation. Experience from the telecommunication and energy utilities. *Utilities Policy*, 24-37.

Pujolle, G., Chaouchi, H., & GaÃ¯ti, D. (2006). A global architecture for the Wi-family. *Telecommunication Systems, 31*(2), 163-173.

Pujolle, G., Chaouchi, H., & Gati, D. (2006). A global architecture for the Wi-family. *Telecommunication Systems, 31*(2), 163-173.

Pyramid Research (2006). *MVNOs and MVNEs: Analyzing the Viability of Virtual Mobile Players.* February 2006 (ID: PYR1224156) Accessed in August 2006. http://www.marketresearch.com/product/display.asp?productid=1224156

Qi, J. Y., Wu, F., & Shu, H. Y. (2007). AI techniques in telecommunications industry. *Expert Systems*, September 2007. *24*(4), 271-291.

Qu, Q., Milstein, L. B., & Vaman, D. R. (2008). Cognitive radio based multi-user resource allocation in mobile ad hoc networks ssing multi-carrier CDMA modulation. *IEEE Journal on Selected Areas in Communications, 26*(1), 70-82.

Quinn, R. E., & Cameron, K. (1983). Organizational life cycles and shifting criteria of effectiveness: some preliminary evidence. *Management Science, 29*,(1), 33-51.

Raffish, N. (1991). How much does that product really cost? *Management Accounting* (US), March, 36-39.

Rafii, F., & Kampas, P. J. (2002). How to identify your enemies before they destroy you. *Harvard Business Review, 80*(11), 115–123.

Rainge, E., & Perry, R. (2004). *Determining the return on investment for automated IP address management.* Whitepaper, IDC. Retrieved May 5, 2007, from http://www.networkworldpartners.com/lucent/pdf/LucentVitalQIPROIWP_Final.pdf

Rajagopalan, B. Electricity over IP, IETF, RFC 3251 (1.4.2002)

Rajagopalan, B., Luciani, J. and Awduche, D. IP over Optical Networks: A Framework, IETF, RFC 3717 (2004)

Rajbanshi, R., Wyglinski, A. M., & Minden, G. J. (2006). Adaptive-mode peak-to-average power ratio reduction algorithm for OFDM-based cognitive radio. *IEEE 64th Vehicular Technology Conference, 1*, 1-5.

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2003). Two-Level Forwarding Pointers-Based Cache Scheme for Location management in PCS Networks. *The 46Th IEEE Midwest Symposium On Circuits and Systems*, Cairo, 226C, Dec. 2003.

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2004). *Two-Level Profile-Based Scheme for Location Update and Paging in PCS Networks.* In *Scientific Bulletin of Ain-Shams University, 39*(4).

Ramadan, S. M., El-Sherbini, A., Marie, M., & Zaki, M. (2007-A). An Integrated Model to Design PCS Networks. *Wireless Personal Communications*, DOI : 10.1007/s11277-006-9126-6, *41*, 19-41.

Ramamurthy, B., Iness, J., & Mukherjee, B. (1998). Optimal amplifier placements in a multiwavelength optical LAN/MAN: the unequally powered wavelength case. *IEEE/ACM Transactions on Networking, 6*(6), 757-767.

Ramamurti, R. (2000). Risks and Rewards in the Globalization of Telecommunications in Emerging Economies. *Journal Of World Business, 35*(2), 149 - 170.

Ramamurti, R., & Sarathy, R. (1997). Deregulation and globalization of airlines. *International Trade Journal, 11*(3), 389-432.

Raman, L. (1998, March). OSI Systems and Network Management. *IEEE Personal Communications, 36*, 46-53.

Ramana, R. K., Singh, S., & Varghese, G. (2007). On scalable attack detection in the network", *IEEE/ACM Transactions on Networking*, *15*, I(1), 14-25. Department of Computer Science, University of California San Diego, La Jolla, CA.

Ramjee, R., Kurose, J., Towsley, D., & Schulzrinne, H. (1994). Adaptive *playout* mechanisms for packetized audio applications in wide-area networks. In INFOCOM (13th IEEE Proceedings), *Networking for Global Communications, 2*, 680-688. Toronto, Canada.

Ranganathan, A., & Campbell, R. H. (2002). Advertising in a Pervasive Computing Environment. *2nd international* *workshop on mobile commerce*. New York, NY: ACM Press.

Rao, B., Angelov, B., & Nov, O. (2006). Fusion of Disruptive Technologies: Lessons from the Skype Case. *European Management Journal, 24*(2-3), 174-188.

Rao, S., & Troshani, I. (2006). *AMC-ER Preliminary Research Report: Understanding Market and Mobile Content Selection for Distribution*. Adelaide, South Australia: Australian Mobile Content – Export Research Initiative, m.Net Corporation Ltd.

Rao, S., & Troshani, I. (2007). A conceptual framework and propositions for the acceptance of mobile services. *Journal of Theoretical and Applied Electronic Commerce Research 2*(2), 61-73.

Rao, S., & Troshani, I. (2007). *AMC-ER Challenges and Rewards: Experiences of Australian Content Exporters*. Adelaide, South Australia: Australian Mobile Content – Export Research Initiative, m.Net Corporation Ltd.

Rao, S., & Troshani, I. (2007). *AMC-ER Initial Selection of Content, Markets, and Distribution Channels for the Australian Mobile Content – Export Research Initiative*. Adelaide, South Australia: m.Net Corporation Ltd.

Rappaport, T. S. (2002). *Wireless Communications Principles and Practice Systems.* India, Delhi, Pearson Education, Inc. Asia. 2002.

Rash, W. (2005). Can only you hear me now? Technologies That Can Secure Voice over IP. *Federal Computer Week, 19*(34), 29.

Rashid, Z. A., Sambasivan, M., & Rahman, A. A. (2004). The influence of organizational culture on attitudes toward organizational change. *The Leadership & Organization Development Journal, 25*(2), 161-179.

Rask, M., & Dholakia, N. (2001). Next to the customer's hart and wallet: Frameworks for exploring the emerging m-commerce arena. *AMA Winter Marketing Educator's Conference, 12*, 372-378.

Ratnasamy, S., Shenker, S., & Stoica, I. (2002). Routing algorithms for DHTs: some open questions. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS02).* Cambridge, MA.

Rautela, D., Markwardt, G., & Kamel, A. (2002). Combining Open APIs (Parlay/JAIN) & Software Agents for Next Generation Mobile Services. *ISMIS02 Workshop.*

Raven Industries. (2008). http://www.ravenprecision.com/Manuals/pdf/016-0171/147A.pdf. (Verified January 23, 2008)

Ray, B., & Abbi, R. (2004). RFC 3728 Definition of managed Objects for Very High Speed Digital Subscriber Lines (VDSL).

Raz, L., Sarit, K., Wilkenfeld, J., & Barry, J. (2007). An *Automated Agent for Bilateral Negotiation with Bounded Rational Agents with Incomplete Information*.

Reagan, J. (2002). The difficult world of predicting telecommunication innovations: Factors affecting adoption. In C. Lin & D. Atkin (Eds.), *Communication technology and society* (pp. 65-90). Cresskill, NJ: Hampton Press.

Reardon, M. (2006a). Qwest CEO supports tiered Internet. *c/net NEWS.Com*, March 16, 2006**.** Retrieved March 16, 2006 from http://news.com.com/Qwest+CEO+supports+tiered+Internet/2100-1034_3-6050109.html.

Reardon, M. (2006b). FCC approves AT&T-BellSouth merger. *c/net NEWS.com*, December 29, 2006. Retrieved May 15, 2008 from http://www.news.com/FCC-approves-ATT-BellSouth-merger/2100-1036_3-6146369.html.

Rec. G.107 (2003). The E-Model, a computational model for use in transmission planning. *ITU-T Recommendation.*

Rec. G.113 (2001). Transmission impairments due to speech processing. *ITU-T Recommendation.*

Rec. G.114 (2000). One-way transmission time. *ITU-T Recommendation.*

Rec. G.723.1 (2006). Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. *ITU-T Recommendation.*

Rec. G.729 (2007). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). *ITU-T Recommendation.*

Rec. H.323 (2006). Packet-based multimedia communications systems. *ITU-T Recommendation.*

Rec. P.800 (1996). Methods for objective and subjective assessment of quality. Rec *ITU-T Recommendation.*

Register. (2005, 14 February 2005). *WiMAX turns the screw on 3G., The Register,* . Retrieved, from the World Wide Web: http://www.theregister.co.uk/2005/02/14/wimax_versus_3g/print.html.

Regli, B. J. W. (1997). *Wireless: Strategically Liberalizing the Telecommunications Market.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Reichheld, F. F., & Teal, T. (1996). *The Loyalty Effect: The Hidden Force behind Growth, Profits and Lasting Value*. Boston, MA: Harvard Business School Press.

Reid, N. P. (2001). *Broadband Fixed Wireless Networks*. New York, NY: McGrawHill/Osborne.

Rekhter, Y., Li, T., & Hares, S. (2006). *A Border Gateway Protocol 4 (BGP-4)*, IETF, RFC 4271.

Renkema, T. (1999). The IT value quest: how to capture the business value of IT-based infrastructure. John Wiley & Sons, Ltd.

Rennie, M. (1993). Global Competitiveness: Born Global. *The McKinsey Quarterly, 4*, 45-52.

Resende, M. (2000). Regulatory regimes and efficiency in US local telephony. *Oxford Economic Papers*, *52*, 447-470.

Reynolds, T., & Gutman, J. (1988). Laddering Theory, Method, Analysis, and Interpretation, *Journal of Advertising Research*. *28*, 11-29.

RFC 3261 (2002). SIP: Session Initiation Protocol. *IETF Standards Track.*

RHK Inc (2003). *Asia-Pacific Capex to Decline, Traffic to Climb.* Annual Forecast Report, Asia-Pacific Telecom Dynamics.

Richard. S. (1997). Combating the churn phenomenon. *Telecommunications*.

Rigney, C., Rubens, A., Simpson, W., & Willens, S. (1997, January). *RFC 2058 Remote Authentication Dial In User Service (RADIUS).*

Riordan, M. H. (1992). Regulation and preemptive technology adoption. *RAND Journal of Economics, 23*(3), 334-349.

Rish, I. (2001). An empirical study of the naive Bayes classifier. *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 41-46). Seattle, USA.

Rix, A. W. (2003). Comparison between subjective listening quality and P.862 PESQ score. *In Proceedings of Measurement of Speech and Audio Quality in Networks (MESAQIN 2003)* (pp. 17-25).

Roberts, J. (1998). *Multinational Business Service Firms: The Development of Multinational Organisational Structures in the UK Business Services Sector.* Aldershot, England: Ashgate Publishing.

Roberts, J. (1999). The Internationalisation of Business Service Firms: A Stages Approach. *Service Industries Journal, 19*(4), 68-88.

Robertson, S., & Robertson, J. (2006). *Mastering the Requirements Process,* 2nd edn. Boston, MA.: Addison-Wesley.

Rogers, E. M. (1962). *Diffusion of Innovation*. New York: The Free Press.

Rohs, M., & Gfeller, B. (2004). Using camera-equipped mobile phones for interacting with realworld objects. In A. Ferscha, H. Hoertner, G. Kotsis (Eds.), *Advances in Pervasive Computing* (pp. 265-271). Vienna, Austria: Austrian Computer Society.

Rondeau, T. W., Rieser, C. J., Gallagher, T. M., & Bostian, C. W. (2004). Online modelling of wireless channels with Hidden Markov Models and channel impulse responses for cognitive radios. *IEEE MTT-S Digest*, 739-742.

Rooney, T. (2007). *Best practices for next-generation IP address management.* White paper. BT Diamond

IP. Retrieved October 24, 2007, from, http://www.pmi.it/file/whitepaper/000133.pdf

Roper-Lowe G, & Sharp J. (1990). Analytic hierarchy process and its application to an information technology decision. *Journal of Operational Research Society, 42*(1), 49-59.

Ros, F., & Ruiz, P. (2006). *Extensible MANET autoconfiguration protocol (EMAP).* IETF draft, March 2006, draft-ros-autoconf-emap-02.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-ros-autoconf-emap-02.

Roseman D. (2003). *The Digital Divide and The Competitive Behavior of Internet Backbone Providers – A Way Forward*. Roseman Associates.

Rosenberg, J., Qiu, L., & Schulzrinne, H. (2000). Integrating packet FEC into adaptive voice *playout* buffer algorithms on the Internet. *In Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000)* (pp. 1705 - 1714), Tel Aviv, Israel.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). *SIP: Session initiation protocol*. Request for Comments (Standards Track), 3261, Internet Engineering Task Force.

Rosenschein, Z. (1989). Negotiation and task sharing among autonomous agents in cooperative domains. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 912-917), Morgan Kaufmann, San Mateo, CA.

Rosman, A., & Rubel, P. G. (1995). *The Tapestry of Culture: an introduction anthropology,* . , New York: McGraw-Hill Inc.

Roto, V., & Oulasvirta, A. (2005). Need for Non-Visual Feedback with Long Response Times in Mobile HCI. *Proceedings of the 14th International World Wide Web Conference*. New York, NY: ACM.

Rousseau, D. (1990). Quantitative assessment of organizational culture: The case for multiple measures. In B. Schneider (Ed.), *Frontiers in Industrial and Organizational Psychology* (pp. 153-193). San Francisco, CA: Jossey-Bass.

Rubino, G., Varela, M., & Bonnin, J.-M. (2006). Controlling multimedia QoS in the future home network using the PSQA metric. *The Computer Journal, 49*(2), 137.
Schulzrinne, H. G., & Rosenberg, J. D. (1998). The Session Initiation Protocol: Providing advanced telephony services across the Internet. *Bell Labs Technical Journal, 3*(4), 144.

Ruef, M., & Scott, W. R. (1998). A multidimensional model of organizational legitimacy: hospital survival in changing institutional environments. *Administrative Science Quarterly*, *43*, 877–904.

Ruffino, S., & Stupar, P. (2006). *Automatic configuration of IPv6 addresses for MANET with multiple gateways (AMG).* IETF draft, June 2006, draft-ruffino-manet-autoconf-multigw-03.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-ruffino-manet-autoconf-multigw-03

Ruggie, J. (1993). Territoriality and beyond: Problematizing modernity in international relations. *International Organisation*, *47*, 139-174.

Ruggles, R. (1998). The state of the notion: knowledge management in practices. *California Management Review*, 40(3), 80-9.

Rugman, A. M. (2003). Regional strategy and the demise of globalization. *Journal of International Managment, 9*(4), 409-417.

Rugman, A. M., & Girod, S. (2003). Retail Multinationals and Globalization: The Evidence is Regional. *European Managment Journal, 21*(1), 24-37.

Russell, S., & Norvig, P. (2001). Artificial Intelligence a modern approach. Prentice Hall series in AI, India.

Rust, R. T., & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing*, *69*, 193-215.

Ryan, B. (2004). *Finance and accounting for business.* London: Thomson Learning.

Saarikoski, V. (2006). *The Odyssey of the Mobile Inter*net. Doctoral Dissertation. University of Oulu, Finland.

Saaty, T. L., & Vargas, L. G. (1994). Decision making in economic, political, social, and technological environments with the analytic hierarchy process. Pittsburgh: RWS Publications.

Sabat, H. K. (2002). The evolving mobile wireless value chain and market structure. *Telecommunications Policy, 26*, 505-535.

Sackmann, S. (2007). Personalization and Privacy in Ubiquitous Computing – Resolving the Conflict by Legally Binding Commitments. *9th IEEE International Conference on E-Commerce Technology*. Washington, DC: IEEE Computer Society.

Sadri G., & Lees B., (2001). Developing corporate culture as a competitive advantage, *Journal of Management Development, 20*,, 10, pp. 853-859

Saha, D., Mukherjee, A., & Bhattacharya, P. S. (2000). A simple heuristic for assignment of cells to switches in a PCS network. *Wireless personal Communications, 12*(3), 209-224.

Sahin, M. E., Ahmed, S., & Arslan, H. (2007). The roles of Ultra Wideband in cognitive networks. *IEEE International Conference on Ultra-Wideband, 1*, 247-252.

Salahaldin, L. (2007). A Numerical Algorithm for Decision-making in Sustainable Transport Projects Investment. *Proceedings of the 11th Annual International Conference on Real options*, Berkeley, CA, June 2007.

Salahaldin, L., & Granger, T. (2005). Investing in Sustainable Transport to Relieve Air Pollution under Population-Growth Uncertainty. *Proceedings of the 9th Annual International Conference on Real options*, Paris, June 2005.

Sallé, M. (2004). *IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing*. Trusted Systems Laboratory, HP Laboratories Palo Alto.

Salleh, Y., & Goh, W. K. (2002). Managing human resources toward achieving knowledge management. *Journal of Knowledge Management, 6*(5), 457-468.

Salmi, A. (1995). *Institutionally changing business networks: An analysis of a Finnish company's operations in exporting to the Soviet Unions, Russia, and the Baltic States*. Doctoral dissertation. Helsinki School of Economics and Business Administration.

Saloner, G., & Shepard, S. (1995). Adoption of Technologies with Network Effects: an Empirical Examination of the Adoption of Automated Teller Machines. *RAND Journal of Economics*, *26*, 479-501.

Saltzer, J. H., Reed, D. P., & Clark, D. D. (1984). End-To-End Arguments in System Design. *ACM Transactions in Computer Systems*, *2*(4), 277-288.

Samhat, A., Skehill, R., & Altman, Z. (2007). Automated Troubleshooting in WLAN Networks. *Proceedings 16th IST Mobile and Wireless Communication Summit*. Budapest, Hungary.

Sanchez, P., Chaminade, C., & Olea, M. (2000). Management of intangibles: an attempt to build a theory, *Journal of Intellectual Capital, 1*(4), 312-327.

Sanders, N. R. (1995). Managing the forecasting function. *Journal of Industrial Management & Data Systems, 95*(4), 12 – 18.

Sangwan, S., & Pau, L.-F. (2005). Diffusion of mobile terminals in China. *European Management Journal, 23*(6), 674-681.

Sanso, B., & Soriano, P. (1999). *Telecommunications Network Planning*. Kluwer Academic Publishing.

Sappington, D. E. M., & Weisman D. L. (1996). Potential pitfalls in empirical investigations of the effects of incentive regulation plans in the telecommunications industry. *Information Economics and Policy, 8*(2), 125-140.

Sappington, D., & Weisman, D. (1996). *Designing incentive regulation for the telecommunications industry.* MIT Press, Cambridge, MA.

Saraydar, C. U., Mandayam, N. B., & Goodman, D. J. (2002). Efficient power control via pricing in wireless data networks. *IEEE Transactions on Communications, 50*(2), 291-303.

Sarkar, M. B., Cavusgil, S. T., & Aulakh, P. S. (1999). International Expansion of Telecommunication Carriers: The Influence of Market Structure, Network Characteristics, and Entry Imperfections. *Journal of International Business Studies*, *30*(2), 361-382.

Sarker, S., & Wells, J. (2003). Understanding Mobile Wireless Device Use and Adoption,. *Communications of the ACM, 46 (12). 46*(12).

Saroit Ismail, I. A. (2000). Bandwidth problems in high-speed networks. *IBM J. Res. Develop.*, *44*(6), 919-938.

Sarrocco C. (2002). *Improving IP connectivity in the Least Developed Countries*, Retrieved May 17, 2004, from http://www.itu.int/osg/spu/ni/ipdc/index.html.

Sascha, S., Holger, F., Matthias, T., & Wolfgang, E. (2004). Traffic view: Traffic data dissemination using car-to-car communication. *ACM Mobile Computing and Communications Review (MC2R)*, *8*(3), 6-19.

Sascha, S., Holger, F., Matthias, T., & Wolfgang, E. (2006). Vehicular ad-hoc networks: Single-hop broadcast is not enough. In H. Rohling (Ed.), *Third International Workshop on Intelligent Transportation* (pp. 49-54), IEEE: Germany.

Sassen, S. (1991). *The global city: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Sato, K.-I. (2007). Recent developments in and challenges of photonic networking technologies. *IEICE Trans. Commun.*, *E90–B*(3), 454-467.

Satti, M. M., & Garner, B. J. (2001). Information security on internet enterprise managed intrusion detection system (EMIDS). *IEEE International Multi-Topic Conference,* (pp. 234 -238). Sydney, NSW, Australia: Macquarie Corporate Telecommunication.

Sattid, M. M., Garner, B. J., & Nagrial, M.H. (2002). Information security standards for e-businesses. In E. Callaway (Ed.), *The Eighth International Conference on Communication Systems* (pp. 641-645). IEEE: USA.

Savolaine, C. (2001). QoS/VoIP overview. *In Proceedings of IEEE Communications Quality & Reliability International Workshop (CQR 2001).*

Sawaragi, Y., Nakayama, H., & Tanino, T. (1985) *Theory of Multipleobjective Optimization*. New York, NY: Academic Press

Saydam, T., & Sirsikar, R. (1998). Design of CMIP-SNMPv2 Proxy Gateway for Network Management Interoperability. *Journal of Networks and Systems Management*, *6*(2).

ScanR (2008). Retrieved on February 15 from http://www.scanr.com

Scapens, R. W. (1994). Never mind the gap: Towards an institutional perspective on management accounting practice. *Management Accounting Research*, *5*, 301-321.

Schatz, A. (2007). FCC Deregulates Verizon's Big-Business Market. *The Wall Street Journal Online Edition*, Retrieved March 21, 2007.

Schein, E. (1990). Organizational culture. *American Psychologist*, *45*, 109–119.

Schein, E. (1992). *Organizational culture and leadership*. A dynamic view (2nd Edition), San Francisco, Jossey – Bass

Schein E. H. (1983). Organizational culture. *Organizational Dynamics 12*, 13– 28.

Schein, E. H. (1985). How culture forms, develops and changes. In Ralph, H.K., Mary, J.S. & Roy, S. (Eds), *Gaining Control of the Corporate Culture*. Jossey-Bass, San Francisco, CA.

Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications, *1st International Workshop on Mobile Computing Systems and Applications*. Washington, DC: IEEE Computer Society.

Schiller, D. (1999). *Digital capitalism: Networking the global market system*. Cambridge, MA: MIT Press.

Schmalensee, R. (2004). *Sunk Costs and Antitrust Barriers to Entry.* MIT Sloan School of Management Working Paper 4457-04.

Schmitz, W. New telecommunication applications and their impact on network platforms - Future of information technology, Proc. 50th anniversary Informationstechnische Gesellschaft ITG, Frankfurt, Germany (2004) 61-68

Schniederjans, M., Hamaker, J., & Schniederjans ,A. (2005). *Information technology investment – decision-making methodology*. Singapore: World Scientific Publishing.

Schoder, D. (2000). Forecasting the success of telecommunication services in the presence of network effects. *Information Economics and Policy, 12*(2), 181-200.

Schoenwaelder, J., Quittek, J., & Kappler, C. (2000). Building Distributed Management Applications with the IETF Script MIB. *IEEE Journal on Selected Areas in Communications*, *18*(5), 702-714.

Schönwälder, J., Pras, A., & Martin-Flatin, J. P. (2003). On the Future of Internet Management Technologies. *IEEE Communications Magazine*, *41*(10), 90-97.

Schowengerdt, R. A. (1997). *Remote Sensing: Models and Methods for Image Processing.* New York, NY: Academic Press.

Schroeder-Bernhardi, J. Analysis of the communication and traffic in P2P networks including Web caches, Diploma Thesis, Report KOM-D-260, Darmstadt Univ. of Tech. (2006)

Schuler, D., Rehbein, K., & Cramer, R. (2002). Pursuing strategic advantage through political means: A multivariate approach. *Academy of Management Journal*, *45*, 659–672.

Schuler, R. (2000). The internationalization of human resource management, *Journal of International Management, 6*(3), pp. 239-60.

Schuler, R., & Jackson, S. (2001). HR issues and activities in mergers and acquisitions, *European Journal of Management, 19*(3), 239-53.

Schuler, R., & Jackson, S. (2001). HR issues and activities in mergers and acquisitions. *European Journal of Management, 19*(3), 239-53.

Schulzrinne, H. G. (1999). Service for Telecom Version 2. *IEEE Internet Computing, 3*(3), 40-43.

Schulzrinne, H., & Rosenberg, J. (1999). Internet telephony: Architecture and protocols - an IETF perspective. *Computer Networks*, *31*(3), 237-255.

Schulzrinne, H., & Rosenberg, J. (2000). The Session Initiation Protocol: Internet Centric Signaling. *IEEE Communications Magazine*, *38*(10), 134-141.

Schumacker, R. E., & Lomax R. G. (2004). *A Beginner's Guide to Structural Equation Modeling.* 2nd Edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Schwiderski, S., Heiko, G. K. (2002). *Secure M-Commerce*.

Scoblete, G. (2007). Competitors swarm to stranded SunRocket subscribers. *TWICE: This Week in Consumer Electronics*, *22*(16), 41-41.

Scott, W. R. (1995). *Institutions and organizations.* Thousands Oaks

Scutari, G., Barbarossa, S., & Ludovici, D. (2003). Cooperation diversity in multihop wireless networks using opportunistic driven multiple access. *4th IEEE Workshop on Signal Processing Advances in Wireless Communications*, *1*, 170-174.

Scutari, G., Barbarossa, S., & Palomar, D. P. (2006). Potential games: a framework for vector power control problems with coupled constraints. *IEEE International Conference on Acoustics, Speech and Signal Processing, 4*, 241-244.

Searcey, D. (2007). Cable's Picture Gets Fuzzier. *The Wall Street Journal*, November 8, 2007. p. B3.

Seeling, P., Fitzek, F. and Reisslein, M. Video Traces for Network Performance Evaluation, Springer (2006)

Semenov , A. B., Strizhakov, S. K., & Suncheley, I. R. (2002). *Structured Cable Systems* (1st ed.): Springer.

Sendonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity. Part I. System description. *IEEE Transactions on Communications, 5*(11), 1927-1938.

Sendonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity. Part I. System description. *IEEE Transactions on Communications, 5*(11), 1927-1938.

Serif, T., Ghinea, G., Stergioulas, L., Chen, S., Tiropanis, T., & Tsekeridou, S. (2007). Satellite-based delivery of educational content to geographically isolated communities: a service based approach. *Personal and Ubiquitous Computing.*

Settles, C. (2007, June). Making the Case for AMI and ROI: Taking the automated meter reading (AMR) to the next great level. *Municipal Wireless Snapshot Report.*

Sexton, M., & Reid, A. (1997). *Broadband Networking: ATM, SDH, and SONET.* Artech House, Norwood, MA.

Shaffer, R. A. (2000). M-Commerce: Online Selling's Wireless Future.

Shane, S., Venkataraman, S., & MacMillian, I. (1995). Cultural differences in innovation championing strategies, *Journal of Management, 21*, 931-52.

Shanks, G., Rouse, A., & Arnott, D. (1993). *A review of approaches to research and scholarship in information systems.* Paper presented at the Fourth Australian Conference in Information Systems, Brisbane, Queensland.

Shannon, V. (2006). Group plans to build China-U.S. fiber optic link. *New York Times* December 19. http://www.nytimes.com/2006/12/19/technology/19cable.html?pagewanted=all

Shapiro, C., & Varian, H. R. (1999). *Information rules: a strategic guide to the network economy.* Boston, Mass.: Harvard Business School Press.

Sharp, A., & Kurkovsky, S. (2007). VOICE: a Framework for Speech-Based Mobile Systems. *2nd IEEE International Symposium on Pervasive Computing and Ad Hoc Communications.* Washington, DC: IEEE Computer Society.

Sharp, B., & Sharp, A. (1997). Loyalty Programs and Their Impact on Repeat-purchase Loyalty Patterns. *International Journal of Research in Marketing, 14*, 473-486.

Sharp, H., Rogers, Y., & Preece, J. (2007). *Interaction Design: Beyond Human-Computer Interaction 2nd Ed.* London, UK: John Wiley & Sons.

Shastry, N., & Adve, R.S. (2006). Stimulating cooperative diversity in wireless ad hoc networks through pricing. *IEEE International Conference on Communications, 8*, 3747-3752).

Shell, G. R. (2004). Make the Rules Or Your Rivals Will. Crown Business: New York.

Shell, G. R., & Yao, D. (2000). *Strategy in a World with Uncertain Rules*, working paper. Philadelphia, University of Pennsylvania: The Wharton School.

Shepherd, A., & Stammers, R. B. (2005). Task analysis. In J. R. Wilson, & N. Corlett, (Eds.). *Evaluation of Human Work*. 3rd edn. London, UK: Taylor & Francis.

Sher, M., & Magedanz, T. (2007). Protecting IP Multimedia Subsystem (IMS) server delivery platform from Time Independent Attacks. *3rd International Symposium on Information Assurance and Security*, (pp. 171-176). Manchester, UK: Technical University, Berlin, Germany.

Sheridan, J. E. (1992). Organizational culture and employee retention, *Academy of Management Journal, 35*, 1036-56.

Shibin S., Ng, J. K.-Y., & Bihai, T. (2004). Some results on the self-similarity property in communication networks, *IEEE Transactions on Communications, 52,* I(10), 1636-1642. Dept. of Risk Manage. & Insurance, Zhongshan Univ., Guangzhou, China.

Shillinglaw, G. (1982). *Managerial cost accounting.* Richard D. Irwin Inc.

Shilun, C., & Zhen, Y. (2007). Energy-efficient power control game for cognitive radio systems. *8th ACIS International Conference on Software Engineering,* *Artificial Intelligence, Networking, and Parallel/Distributed Computing, 1*, 526-530.

Shin, D. (2007). A critique of Korean national information strategy: case of national information infrastructures. *Government Information Quarterly, 24*(3), 624–645.

Shin, D. H. (2007). A study of mobile number portability effects in the United States. *Telematics and Informatics, 24*(1), 1-14.

Shin, D. H., & Bartolacci, M. (2007). A study of MVNO diffusion and market structure in the EU, US, Hong Kong, and Singapore. *Telematics & Informatics, 24*, 86-100.

Shin, D., Kim, J., Ryu, S., Oh, D., Lee, J., & Kang, M. (2006). Scenario Decomposition Based Analysis of Next Generation Mobile Services. *Proceedings of the International Conference on Computational Science, (2)* 977-984.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer interaction*, 3rd edn. Reading, MA: Addison Wesley Longman.

Shrum, W., & Wuthnow, R. (1988). 'Reputational status of organizations in technical systems'.*American Journal of Sociology, 93*, 882–912.

Shu, H. Y., & Qi, J. Y. (2004). *Telecommunication Customer Life Cycle Management*. Beijing University of Posts and Telecommunications Press.

Shwiff, K. (2007). Legal Costs Widen Vonage's Loss. *The Wall Street Journal Online Edition*, Retrieved November 8, 2007.

Shy, O. (2002). *The economics of network industries*. Cambridge: Cambridge University Press.

Sidak, G. (2007). What Is the Network Neutrality Debate Really About? *International Journal of Communications, 1*(1), 377-388.

Siegemund, F., Floerkemeier, C., & Vogt, H. (2005). The value of handhelds in smart environments. *Personal and Ubiquitous Computing, 9*(2), 69-80.

Sierra, C. N, Jennings, R., Noriega, P., Parsons, S. (1998). A framework for argumentation-based negotiation. *Proceedings of ATAL 1997, 1365*, 177-192. Berlin, Germany: Springer-Verlag.

Sigurdson, J. (2001). WAP OFF - Origin, Failure and Future. *Prepared for the Japanese-European Technology Studies (JETS)*, October 9, 2001, University of Edinburgh.

Silos, I. (1999). Employee involvement – a component of total quality management. *Production & Inventory Management Journal, 40*(1), 56-65.

Silvester, J., & Anderson, N. R. (1999). Organizational culture change. *Journal of Occupational & Organizational Psychology, 72*(1), 1-24.

Simeone, O., Gambini, J., Bar-Ness, Y., & Spagnolini, U. (2007). Cooperation and Cognitive Radio. *IEEE International Conference on Communications,.1*, 6511-6515.

Simoneau, P. (1999). *SNMP Network Management*. New York: McGraw-Hill Companies.

Simpson, N. (1999). Managing the use of style guides in an organisational setting: Practical lessons in ensuring UI consistency. *Interacting with Computers, 11*(3), 323-351.

Simpson, P. M., Siguaw, J. A., & Baker, T. L. (2001). Value Creation Supplier Behaviors and Their Impact on Reseller-Perceived Value. *Industrial Marketing Management, 30*, 119-134.

Singhi, M., & Long, H. (1998). New undersea cable developments and satellite services: Toward complementary coexistence in the 21st Century. *Proceedings of the 20th Pacific Telecommunications Conference* (pp. 566-569). Honolulu: Pacific Telecommunications Council.

Sinkula, J.M. (1994). Market-information processing and organizational learning, *Journal of Marketing, 58*, 1, pp. 35-45.

Siregar, J., Tagaki, H., & Zhang, Y. (2002). Optimal Wavelength Converter Placement in Optical Networks by Genetic Algorithm. *IEICE Trans. Commun., E86*(B(6)), 1075-1082.

Sirel, E., & Wavermann, L. (2000). 3G Mobile: The Push to Wireless Data across Europe. *Business Strategy Review, 11*(2), 67-71.

Sivananthiran, A., & Venkata Ratnam, C. S. (2004). *Social dialogue at enterprise level: Successful experients*. International Labour Organization, available at: www.ilo.org/public/english/dialogue/ifpdial/downloads/papers/social_dia_entlevel.pdf

Skaanning, C. (2000). A Knowledge Acquisition Tool for Bayesian-Network Troubleshooters. *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence* (pp. 549-557). Stanford, USA.

Skaanning, C., Jensen, F. V., & Kjaerulff, U. (2000b). Printer troubleshooting using Bayesian Networks. *Lecture Notes in Artificial Intelligence*, *1821*, 367-379.

Skog, B. (2000). The mobile phone as symbolic capital in teenage cultures. (in Norwegian). In *R. Ling and K. Thrane (Eds.), Social consequences of mobile telephony* (in Norwegian). *Proceedings from a seminar on society, children, and mobile telephony*, Telenor FoU - Notat nr. 38/2000.

Skog, B. (2002). Mobiles and the Norwegian teen: identity, gender and class. In J. E. Katz and M. Aakhus (Eds.), *Perpetual contact*. New York: Cambridge University Press.

Skorin-Kapov, N. (2007). A new objective criterion and rounding techniques for determining virtual topologies in optical networks. *IEEE Communication Letters, 11*(6), 540-542.

Skorin-Kapov, N. (2006). Multicast routing and wavelength assignment in WDM networks: a bin packing approach. *Journal of Optical Networking, 5*(4), 266-279.

Skorin-Kapov, N. (2007). Routing and wavelength assignment in optical networks using bin packing based algorithms. *European Journal of Operational Research, 177*(2), 1167-1179.

Skyhook. (2008). *Wi-Fi Positioning*. Retrieved January 30, 2008, from http://www.skyhookwireless.com/

Skype. (2007). *Skype Technical FAQ*. Retrieved October 21, 2007 from http://www.skype.com/help/faq/technical.html

Slot, J. (2007). Changing user roles in ICT developments: The case of digital television. *Telematics & Infomatics, 24*, 303-314.

Smith, A. M., & Fischbacher, M. (2005). New service development: a stakeholder perspective. *European Journal of Marketing*, *39*(9/10), 1025-1048.

Smith, D. (2003). *Digital transmission systems*. London: Springer.

Smith, K. G., Grimm, C. M., & Gannon, M. J. (1992). *Dynamics of competitive strategy*. London: Sage Publications.

Smith, R. G. (1980). The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computer, 29*, 1104-1113.

Soliman, F., & Spooner, K. (2000). Strategies for implementation knowledge management: Role of human resource management. *Journal of Knowledge Management*, *4*(4), 337-345.

Spiller, P. T. & Cardilli, C. (1999). Towards a property rights approach to communication spectrum. *Yale Journal on Regulation*, *16*(1), 53-84.

Spiller, P. T., & Zelner, B. A. (1997). Product complementarities, capabilities and governance: a dynamic transaction cost perspective. *Industrial and Corporate Change*, *6*(3), 561-594.

Sprint Nextel News Release (2005). *Sprint Nextel Completes Merger*.

Sprint. (2008). *Sprint and Clearwire to Combine WiMAX Businesses, Creating a New Mobile Broadband Company*. Retrieved, from the World Wide Web: http://newsreleases.sprint.com/phoenix.zhtml?c=127149&p=irol-newsArticle_newsroom&ID=1141088

Sreenan, C. J., Chen, J. C., Agrawal, P., & Narendran, B. (2000). Delay reduction techniques for *playout* buffering. *IEEE Transactions on Multimedia, 2*(2), 88-100.

Srinivasa, S., & Jafar, S. A. (2007a). The throughput potential of cognitive radio: A theoretical perspective. *IEEE Communications Magazine, 45*(5), 73-79.

Srinivasa, S., & Jafar, S. A. (2007b). Cognitive radio networks: How much spectrum sharing is optimal? *IEEE Global Communications Conference, 1*, 3149-3153.

Srinivasan, S.S., Andeson, R., & Ponnavolu, K. (2002). Customer Loyalty in E-commerce: an Exploration of Its Antecedents and Consequences. *Journal of Retailing, 78*, 41-50.

Srinivasan, V., Nuggehalli, P., Chiasserini, C.-F., & Rao, R. R. (2005). An analytical approach to the study of cooperation in wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, 4 (2), 722-733.

St. Amant, R., Horton, T. E., & Ritter, F. E. (2007). Model-based evaluation of expert cell phone menu interaction. *ACM Transactions on Computer-Human Interaction, 14*(1).

St. Jean, C. A., & Jabbari, B. (2004). Bayesian game-theoretic modeling of transmit power determination in a self-organizing CDMA wireless network. *IEEE 60th Vehicular Technology Conference, 5*, 3496-3500.

Staehle, B., Binzenhöfer A., Schlosser, D. and Boder, B. Quantifying the influence of network conditions on the service quality experienced by a thin client user, Proc. 14. Conf. on Measurement, Modeling and Evaluation of Comp. and Comm. Systems (MMB), Dortmund, Germany (2008) 173-187

Stallings, W. (1998). *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2*. Indianapolis, Addison-Wesley Professional.

Stallings, W. (2002). *Data and Computer Communications*. New York: Macmillan Publishing Company.

Staple, G. (2004). *Telegeography 2004: Global telecommunications traffic statistics and commentary*. Washington, DC: Telegeography, Inc.

Star, S.L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructures: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1):111-134.

Staubus, G. J. (1971). Activity costing and input-output accounting. In Staubus, G. J. (1988), *Activity costing for decisions: Cost accounting in the decision usefulness framework*. Garland Publishing, Inc.

Steensma, K. H., Marino, L., Weaver, M. K., & Dickson, P. H. (2000). The influence of national culture on the formation of technology alliances by entrepreneurial firms. *Academy of Management Journal, 43*(5), 951-73.

Stefano, C., Heikki, H., Oskari, K. (2000). Performance enhancing proxies for Java2 RMI over slow wireless links. *International Conference on the Practical Application of Java, pp.* 76-89.

Stein, Y(J)., Shashoua, R., Insler, R., Anavi, M. Time Division Multiplexing over IP, IETF, RFC 5087 (2007)

Steinbock, D. (2003). Globalization of wireless value system: from geographic to strategic advantages. *Telecommunications Policy, 27*(3/4), 207-235.

Steinbock, D. (2005). *The Mobile Revolution: The Making of Mobile Services Worldwide*. London: Kogan Page.

Steinder, M., & Sethi, A. S. (2004). Probabilistic Fault Localization in Communication Systems Using Belief Networks. *IEEE/ACM Transactions on Networking, 12*(5), 809-822.

Stephan, E. (2007). A security architecture concept for Vehicular Network Nodes. In C. H. Kam, (Ed.), T*he Sixth International Conference on Information, Communications and Signal Processing.* IEEE: Singapore.

Sterbenz, J., & Touch, J. (2001). *High-speed networking: A systematic approach to high-bandwidth low-latency communication*., New York, US: John Wiley.

Sterling, C., Bernt, P., & Weiss, M. (2006). *Shaping American Communications*. New York, N.Y.: Lawrence Erlbaum Associates.

St-Hilaire, M., Chamberland, S., & Pierre, S. (2006). Uplink UMTS Network Design - an integrated approach. *Computer Networks, 50*, 2747-2761.

Stienstra, M., Baaij, M., & Bosch, F. V. D. (2004). Strategic renewal of Europe's largest telecom operator (1992-2001): from herd behavior towards strategic choice? *European Management Journal*, *22*(3), 273-280.

Stienstra, M., Baaij, M., Van den Bosch, F., & Volberda, H. (2004). Strategic Renewal of Europe's Largest Telecom Operators (1992-2001): From Herd Behaviour Towards Strategic Choice? *European Managment Journal*, *22*(3), 273-280.

Stipp, H. (1999). Convergence now? *The International Journal on Media Management, 1*(1), 10-13.

Stoddard, D., & Jarvenpaa, S. (1995). Business process redesign: Tactics for Managing Radical Change. *Journal of Management Information Systems*, *12*(1),81-107.

Stoica, I. et al., Chord: A scalable P2P lookup protocol, IEEE/ACM Trans. on Networking Vol. 11 (2003) 17-32

Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., & Balakrishnan, H. (2003). Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on Networking*, *11*(1), 17-32.

Strassner, J. (2004). Autonomic networking – Theory and practice. *IEEE/IFIP Network Operations and Management Symposium*.

Straub, D., Loch, K., Evaristo, R., Karahanna, E., & Srite, M. (2002). Toward a theory-based measurement of culture, *Journal of Global Information Management*, *10* (1), 13–23.

Strauss, F. (2003). *libsmi - A Library to Access SMI MIB Information*. From http://www.ibr.cs.tu-bs.de/projects/libsmi/

Strauss, F., & Klie, T. (2003). Towards XML Oriented Internet Management. In *8th IFIP/IEEE International Symposium on Integrated Network Management* (pp. 505-518).

Stuart, T. E. (2000). Interorganizational alliances and the performance of firms: a study of growth and innovation rates in a high-technology industry. *Strategic Management Journal*, *21*(8), 791-811.

Su, A.-J., Choffnes, D.R., Kuzmanovic, A. and Bustamante, F.E. Drafting behind Akamai, ACM SIGCOMM, Pisa, Italy (2006)

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571 – 611.

Sumair, U. R. & Urs, H. (2007). Secure Crash Reporting in Vehicular Ad hoc Networks. In B. Crispo (Ed.), *The Third International Conference on Security and Privacy in Communication Networks ((SecureComm 2007)* (pp. 39-68), IEEE: France.

SUN jNETx OSA Platform at www.sun.com

Sun, C., Zhang, W., & Letaief, K. B. (2007). Cluster-based cooperative spectrum sensing in cognitive radio systems. *IEEE International Conference on Communications, 1*, 2511-2515.

Sun, L., & Ifeachor, E. (2004). New models for perceived voice quality prediction and their applications in *playout* buffer optimization for VoIP networks. *In Proceedings of IEEE International Conference on Communications (ICC 2004),. 3*, 1478-1483.

Sveiby, K.E., Simons, R. (2002). "Collaborative climate and effectiveness of knowledge work-an empirical study", *Journal of Knowledge Management*, 6 No.5, pp.420-33.

Sveiby, K. E. (2000). *What is knowledge management*. Retrieved March 18, 2005, from http://www.sveiby.com.ay/KnowledgeManagement.html.

Swanson, B. (2007). The Coming Exaflood. *The Wall Street Journal Online Edition*, Retrieved January 20, 2007.

Sydow, J., & Windeler, A. (1998). Organizing and evaluating inter firm networks: A structurationist perspective on network processes and effectiveness. *Organizational Science*, *9*(3), 265-284.

Syed-Ikhsan, S. O. S., & Rowland, F. (2004). Benchmarking knowledge management in a public organization in Malaysia. *Benchmarking: An International Journal, 11*(3), 238-266.

Symes, P. (1998). *Video compression: Fundamental compression techniques and an overview of the JPEG and MPEG compression systems.* NY: McGraw-Hill.

Szabó, Cs., Chlamtac, I., & Bedő, E. (2004, January). Design Considerations of Broadband Community Networks. In *Proceedings of the Thirty-Seventh Annual Hawaii International Conference on System Science (HICSS 2004).* Hawaii, USA: Computer Society Press

Szabó, Cs., Farkas, K., & Horváth, Z. (2008). Motivations, Design and Business Models of Wireless Community Networks. To be published in *Mobile Networks and Applications* journal: Springer

Szabó, Cs., Horváth, Z., & Farkas, K. (2007, October). Wireless Community Networks: Motivations, Design and Business Models. In *Proceedings of the Third Annual International Wireless Internet Conference (WICON 2007).* Austin, Texas, USA.

T1A1.7/98-031 (1998). Testing the quality of connections having time varying impairments. Committee contribution *ANSI & AT&T.*

Tabbane, S. (1995). An Alternative Strategy for Location Tracking. *IEEE J. Selected Areas in Comm., 13*(5), 880-892.

Tabbane, S. (1997). Location Management Methods for Third-Generation Mobile System. *IEEE Comm. Magazine.*

Tait, D., Keijzer, J. D., Goerdman, R. (2000). JAIN: A New Approach to Services in Communication Networks. *IEEE Communications Magazine, 38*(1), 94-99.

Takahashi, A., Yoshino, H., & Kitawaki, N. (2004). Perceptual QoS assessment technologies for VoIP. *IEEE Communications Magazine, 42*(7), 28-34.

Tallberg, M., Hämmäinen, H., Töyli, J., Kamppari, S., & Kivi, A. (2007). Impacts of handset bundling on mobile data usage: the case of Finland. *Telecommunications Policy, 31*(10/11), 648-659.

Tam, M., & Tummala, R. (2001). An application of the AHP in vendor selection of a telecommunications system. *Omega, 29,* 171-182.

Tamminen, S., Oulasvitra, A., Toiskallio, K., & Kankainen, A. (2004). Understanding Mobile Contexts. *Personal and Ubiquitous Computing, 8*(2), 135-143.

Tang, A., Lee, J.-W., Huang, J., Chiang, M., & Calderbank, A. R. (2006). Reverse engineering MAC. *4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 1,* 1-11.

Tang, D., & Baker, M. (2000). *Analysis of a Metropolitan-Area Wireless Network*. Wireless Networks, Department of Computer Science, Stanford University.

Tarasewich, P. (2003). Designing Mobile Commerce Applications. *Communications of the ACM, 46*(12), 57-60.

Tarasewich, P., Nickerson, R. C., & Warkentin, M. (2002). Issues in Mobile Commerce. *Communications of the Association for Information Systems, 8,* 41-64.

Tardiff, T., & Taylor, W. E. (1993). *Telephone company performance under alternative forms of regulation in the U.S.* National Economic Research Associates.

Tarn, J. M., Razi, M., Yen, D., & Xu, Z. (2002), Linking ERP and SCM systems. *International Journal of Manufacturing Technology & Management, 4*(5), 420-439.

Tarnay, K. (1991). *Protocol Specification and Testing.* New York: Plenum Press.

Tayal, A., & Patnaik, M. (2004). An address assignment for the automatic configuration of mobile ad hoc networks, *Personal and Ubiquitous Computing, 8*(1), 47-54.

Taylor, A. (2000). IT projects: sink or swim. *The Computer Bulletin,* Jan, 24-26.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research, 6,* 144-176.

Taylor, W. E., Zarkadas, C. J., & Zona J. D. (1992, June). *Incentive regulation and the diffusion of new technology in telecommunications.* Paper presented at Ninth Annual Conference, International Telecommunications Society, Sophia-Antipolis, France.

Teece, D. (2000). Strategies for managing knowledge assets: The role of firm structure and industrial context. *Long Range Planning, 33*(1), 35-54.

Templin, F., Russert, S., & Yi, S. (2007). *MANET autoconfiguration*. IETF draft, March 2007, draft-templin-autoconf-dhcp-07.txt. Retrieved December 20, 2007, from, http://tools.ietf.org/id/draft-templin-autoconf-dhcp-07.

Terpstra, W., Kangasharju, J., Leng, C. and Buchmann, A. BubbleStorm: Resilient, probabilistic and exhaustive P2P search, Proc. ACM SIGCOMM, Kyoto, Japan (2007) 49-60

The Parlay Group official Website, http://www.parlay.org/

The Wireless Philadelphia Executive Committee. (2005, February). *Wireless Philadelphia Business Plan: Wireless Broadband as the Foundation for a Digital City.*

Thermos, P. (2006). *Two attacks against VoIP*. Retrieved October 21, 2007 from http://www.securityfocus.com/infocus/1862/2

Thiagarajan, S., & Somani, A. (2003). Optimal wavelength converter placement in arbitrary topology wavelength-routed networks. *Computer Communications, 26,* 975-985.

Thiel, S. U., Giuliani, P., Ibbetson, L. J., & Lister, D. (2002). An automated UMTS site selection tool. *3G Mobile Communication Technologies* (pp. 69-73).

Thomas, R., Gilbert, H., & Mazziotto, G. (1988). Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network. *In Third Nordic Seminar on Digital Land Mobile Radio Communication*. Copenhagen.

Thompson, A., & Strickland, A. (2008). Dell computer corporation, Available at: http://www.mhhe.com/business/management/thompson/11e/case/dell.html. Retrieved: 20 January 2008.

Tichy, N. M. & Cohen, E. (1997). *The Leadership Engine: How Winning Companies Build Leaders at Every Level*, New York, NY, Harper-Collins,.

Timmers, P. (1998). Business Models for Electronic Markets. *Electronic Markets, 8*(2), 3-8. www.it.iitb.ac.in/~annanda/papers-280905/Secure%20m-commerce%20ECEJ.pdf

TIPHON TS 101329-5 (2001). QoS measurement methodologies. *ETSI Technical Standard: Annex E.*

Tiresias.org (2008) Retrieved 28th Feb 2008, from http://www.tiresias.org/guidelines/.

Tiwana, A. (2000). *The knowledge management toolkit: Practical techniques for building a knowledge management system*. 1st ed., Upper Saddle River, NJ: Prentice-Hall PTR.

Tkachenko, A., Cabric, D., & Brodersen, R. W. (2007). Cyclostationary feature detector experiments using reconfigurable BEE2. *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, .1*, 216-219.

Tongia, R. (2007). Connectivity in emerging regions: the need for improved technology and business models. *IEEE Communications Magazine, 45*(1), 96-103.

Tosic, V., Patel, K., & Pagurek, B. (2002). WSOL - Web Service Offerings Language. *Lecture Notes In Computer Science*, *2512*. Revised Papers from the International Workshop on Web services, E-Business, and the Semantic Web, (pp. 57-67).

Townsley, M., Pignataro, C., Wainner, S., Seely, T. and Young, J. Encapsulation of MPLS over Layer 2 Tunneling Protocol Version 3, IETF, RFC 4817 (2007)

Tran, A., Tucker, R., & Boland, N. (2004). Amplifier placement methods for metropolitan WDM ring networks. *Journal of Lightwave Technology, 22*(11), 2509-2522.

Trice H. M. & Beyer J. M. (1984). *The cultures of work organizations*, Englewood Cliffs, NJ: Prentice - Hall

Trigeorgis, L. (1993). The Nature of Options Interactions and the Valuation of Investments with Multiple Real Options. *Journal of Financial and Quantitative Analysis, 28*(1), 1-20.

Trigeorgis, L. (1996). *Real options: managerial flexibility and strategy in resource allocation.* Cambridge MA: The MIT Press.

Tripsa, M. (1997). Unraveling the Process of Creative Destruction: Complementary Assets and Incumbent Survival in the Typesetter Industry. *Strategic Management Journal*, *18* (1), 119-142.

Tront, J. G., & Marchany, R. C. (2004). Internet Security: Intrusion Detection and Prevention. *37th Annual Hawaii International Conference on System Sciences,* (pp. 188-188). Bradley Dept. of Electrical & Computer Engineering, Virginia Tech., Blacksburg, VA.

Troshani, I., & Rao Hill, S. (2008). A Proposed Framework for Mobile Services Adoption: A Review of Existing Theories, Extensions, and Future Research Directions. In G. C. Karmakar & L. S. Dooley (Eds.), *Mobile Multimedia Communications: Concepts, Applications and Challenges* (pp. 85-108). Hershey, PA, USA: IGI Global.

Troshani, I., & Rao, S. (2007a). The diffusion of mobile services in Australia: an evaluation using stakeholder and transaction cost economics theories. *IADIS International Journal of WWW/Internet, 5*(2), 40-57.

Tselikas, N. D., Dellas, N. L., Koutsoloukas, E., Kapellaki, S., Prezerakos, G. N., & Venieris, I. (2007). Distributed service provision using open APIs-based middleware: "OSA/Parlay vs. JAIN" performance evaluation study. *Journal of Systems and Software* 80(5). 765-777.

Tseng, F. C., Teng, C. I., & Chiang, D. M. (2007). Delivering Superior Customer Perceived Value in the Context of Network Effects. *International Journal of E-Business Research, 3*(1), 41.

Tuggle F.D. & Shaw N.C. (2000). "*The effect of organizational culture on the implementation of knowledge management*", Orlando, FL, Florida Artificial Intelligence Research Symposium (FLAIRS).

Turner, C. (2004). Changing strategic behaviour in European telecommunications: the experience of Colt. *Info, 6*(4), 252-260.

Turney, P. B. B. (1991). *Activity based costing: the performance breakthrough*. London: Kogan Page.

Ubacht, J. (2004). Regulatory practice in telecommunications markets: how to grasp the 'real world of public action'? *Communications & Strategies*, (54, 2nd Quarter), 219-242.

Uchiyama, H., Umebayashi, K, Kamiya, Y., Suzuki, Y., & Fujii, T. (2007). Study of cooperative sensing in cognitive radio based ad-hoc network. *18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1*, 1-5.

Ulrich, D. (1997). Measuring human resources: an overview of practice and a prescription for results, *Human Resource Management*, *36*, 3, pp. 303-20.

Ulset, S. (2002). Mobile virtual network operators: a strategic transaction cost analysis of preliminary experiences. *Telecommunications Policy,* 26(9-10), 537-549.

United States District Court for the District of Columbia (2000). *United States of American v. Bell Atlantic, GTE and Vodafone Airtouch, Motion for Entry of Final Judgment.*

United States, General Accounting Office (USGAO) (2001*). Telecommunications: Characteristics and Competitiveness of the Internet Backbone Market*, October.

University of Minnesota, MINTS: Minnesota Internet traffic studies http://www.dtc.umn.edu/mints/references.html

unstrung.com. (2005). *WiMAX Telecom Eyes Standards*. Retrieved, from the World Wide Web: http://www.unstrung.com/document.asp?doc_id=66413

Uppaluru, P. (2003). Discover the power of voice: Transforming the telephone to open new business models. *Alcatel Telecommunications Review*(3), 329-334.

Usero, B., & Fernández, Z. (2005). *First come, first served: An analysis of pioneer and follower firms's market and non-market actions in the European mobile telephone industry* (Business Economics 05-48-12). Madrid, Spain: Universidad Carlos III de Madrid.

Väänänen-Vainio-Mattila, K., & Ruuska, S. (2000). Designing Mobile Phones and Communicators for Consumers' Needs at Nokia. In E. Bergman (Ed.), *Information Appliances and Beyond: Information Design for Consumer Products.* San Francisco, CA: Morgan Kaufmann.

Vakola M. & Nikolaou I., (2005). Attitudes towards organizational change. What is the role of employees stress and commitment, *Employee Relations, 27, 2,* 2005, pp. 160-174

Valentine, S. (2000). International person-organization fit: The role of national culture. *International Journal of Management, 17*(3), 295-302.

Valetti, T. M. (2001). Spectrum trading. *Telecommunications Policy, 25*(10/11), 655-670.

Van De Kar, E. A. M. (2004). *Designing Mobile Information Services: An Approach for Organisations in a Value Network.* Unpublished Doctoral Dissertation, Delft University of Technology, Delft.

van den Berg, H., Mandjes, M., van de Meent, R., Pras, A., Roijers, F., and Venemans, P., QoS-aware bandwidth provisioning for IP backbone links, Computer Networks 50 (2006) 631-647

Van den Ende, J. (2003). Modes of governance of new service development for mobile networks: A life cycle perspective. *Research Policy*, *32*(8), 1501-1518.

Vanderhorn, N., Balasubramanian, S., Mina, M., Weber, R. J., & Somani, A. K. (2006, March). Light-Trail Testbed for Metro Optical Networks. *Proc. of the 2nd International IEEE/Create-Net Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, TRIDENTCOM* .

Vanhonacker, W. (2000). A better way to crack China. *Harvard Business Review,* July-August, 130-136.

Varshney, U., Snow, A., McGivern, M., & Howard, C. (2002). Voice over IP. *Communications of the ACM, 45*(1), 89-96.

Vaughan-Nichols, S. J. (2004). Achieving wireless broadband with WiMax. *IEEE Computer*, *37*(6), 10- 13.

Venieris, I., & Hussmann, H. (Ed.) (1998). *Intelligent Broadband Networks.* John Wiley & Sons Ltd.

Venieris, I., Zizza, F., & Magedanz, T. (Eds.) (2000). *Object Oriented Software Technologies in Telecommunications.* John Wiley, Chichester.

Venkatesh, V. R., et.al. (2003). Understanding Usability in Mobile Commerce. *Communications of the ACM, 46*(12), 53-56.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science, 46*(2), 186-204.

Verespej, M. (1999). Knowledge management: System or culture? *Industry Week, 248,* 20-21.

Verikoukis, C., Konstas, I., Anastasiadou, D., & Angelidis, P. (2006). Overview on telecommunications regulation framework in south-eastern Europe. *Telecommunication Systems*, *32*(2/3), 209-221.

Verizon (2007). Verizon Continues to Dramatically Raise Broadband Upload Speeds in FiOS Internet Service Area. Verizon News Release, November 20.

Verkasalo, H. (2007). *A Cross-Country Comparison of Mobile Service and Handset Usage*. Licentiate's thesis, Helsinki University of Technology, Networking Laboratory, Finland.

Verkasalo, H. (2007). Contextual Usage-Level Analysis of Mobile Services. *Accepted for publication at The 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS 2007).* August 6-10, 2007 - Philadelphia, PA, USA.

Verkasalo, H. (2007). Empirical Findings on the Mobile Internet and E-Commerce. *Presented at 20th Bled e Conference: eMergence: Merging and Emerging Technologies, Processes, and Institutions.* Bled, Slovenia, June 4 - 6, 2007.

Verkasalo, H. (2007). Empirical Insights on the Evolution of the Finnish Mobile Market. Presented at Conference on Telecommunication Techno-Economics (CTTE) 2007, 14-15 June 2007, Helsinki, Finland.

Verkasalo, H. (2006). Empirical Observations on the Emergence of Mobile Multimedia Services in the U.S. and Europe. In *The 5th International Conference on Mobile and Ubiquitous Multimedia*. December 4-6, 2006, Stanford University, California, 2006.

Verkasalo H. (2005). *Handset-Based Monitoring of Mobile Customer Behavior.* Master's Thesis Series. Networking Laboratory. Department of Electrical and Telecommunications Engineering. Helsinki University of Technology.

Verkasalo, H. (2006). Mobile Data Service Evolution - Empirical Implications from Europe and the USA. In *3rd International CICT Conference*, November 30 - December 1, 2006, Denmark, 2006.

Verkasalo, H., & Hämmäinen, H. (2006). Handset-Based Monitoring of Mobile Subscribers. *Presented at Mobility Roundtable*, 1-2 June, 2006, Helsinki, Finland.

Verkasalo, H., & Hämmäinen, H. (2007). A Handset-Based Platform for Measuring Mobile Service Usage. *INFO: The Journal of Policy, Regulation and Strategy, 9*(1), 2007.

Vesa, J. (2005). *Mobile Services in the Networked Economy*. Hershey, PA, USA: IRM Press

ViaTalk. (2007). *How to set up your ViaTalk broadband phone*. Retrieved November 27, 2007, from http://support.viatalk.com

Visotsky, E., Kuffner, S., & Peterson, R. (2005). On collaborative detection of TV transmissions support of dynamic spectrum sharing. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, *1*, 338-345.

VoIP Mechanic. (2006). How to correct VoIP echo. Retrieved November 26, 2007, from http://www.voip-mechanic.com/phoneandvoicequality.htm

voip-info.org. (2007). ATA. Retrieved October 22, 2007, from http://www.voip-info.org/wiki-ATA

Vu, L., Hauswirth, M., & Aberer, K. (2005). QoS-based Service Selection and Ranking with Trust and Reputation Management. *Infoscience's Technical Report*.

Vuran, M. C., & Akyildiz, I. F. (2007). A-MAC: Adaptive Medium Access Control for next generation wireless terminals. IEEE/ACM Transactions on Networking, *15*(3), 574-587.

W3C (2006). *Mobile Web Best Practices 1.0*. Retrieved 28th Feb 2008, from http://www.w3.org/TR/mobile-bp/

W3C, Web Services Description Language (WSDL) Specification, v 1.1, at http://www.w3.org/TR/wsdl

Waclawsky, J. G. (2007). SIP's future: Complicated and competitive. *Business Communications Review, 37*(2), 22.

Wainwright, D. W., & Waring, T. S. (2007). The application and adaptation of a diffusion of innovation framework for information systems research in NHS general medical practice. *Journal of Information Technology, 22*(1), 44-58.

Wakikawa, R., Malinen, J., Perkins, C., Nilsson, A., & Tuominen, A. (2006). *Global connectivity for IPv6 mobile Ad-hoc networks*. IETF draft, March 2006, draft-wakikawa-manet-globalv6-05.txt. Retrieved December 14, 2007, from http://www.tools.ietf.org/html/draft-wakikawa-manet-globalv6-05

Walcott, S., & Wheeler, J. (2001). Atlanta in the telecommunications age: The fiber-optic information network. *Urban Geography*, *22*, 316-339.

Walker, B. A., & Olson, .J. C. (1991). Means-end chains: Connecting products with sefl, *Journal of Business Research*, *22*(2), 111-118.

Walker, W. R. (1989). *Guidelines for Designing and Evaluating Surface Irrigation Systems*. FAO, 1989: ISBN 92-5-102879-6, http://www.fao.org/docrep/T0231E/t0231e01.htm#preface (Verified January 23, 2008)

Wallsten, S. (2005). *Broadband penetration: An empirical analysis of state and federal policies* (Working Paper 05-12). Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Wallsten, S. (2006). *Broadband and unbundling regulations in OECD countries* (Working Paper 06-16). Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Walter, G.A. (1985). *Culture collisions in mergers and acquisitions*, in Frost, P.J. et al. (Eds), Organizational Culture, Sage, Beverly Hills, CA, pp. 301-14.

Wan, Y. H., Qian, Y. Y., & Li, J. (2005). The Analysis of Customer Value Dimensions' Difference of the Bank's Interacting Channel, *The Second IEEE Conference on Service Systems and Service Management*, (pp. 193-198).

Wang, B., Han, Z., & Liu, K. J. R.(2006). Stackelberg game for distributed resource allocation over multiuser cooperative communications networks. *IEEE Global Communications Conference,1*, 1-5.

Wang, H. L., & Feng Y. Q. (2007). Decision support system facing unstructured decision making based on decision flow. *Journal of Harbin Institute of Technology, 10*, 1638-1641.

Wang, K.-C., & Ramanathan, P. (2005). QoS assurances through class selection and proportional differentiation in wireless networks. *IEEE Journal on Selected Areas in Communications*, *23*(3), 573-584.

Wang, S., & Archer, N. (2007). Electronic marketplace definition and classification: literature review and clarification. *Enterprise Information Systems*, *1*(1), 89-112.

Wang, W., Cui, Y., Peng, T., & Wang, W. (2007). Non-cooperative power control game with exponential pricing for cognitive radio Network. *IEEE 65th Vehicular Technology Conference, 1*, 3125-3129.

Wang, W., Peng, T., & Wang, W. (2007). Optimal power control under interference temperature constraints in cognitive radio network. *IEEE Wireless Communications and Networking Conference, 1,* 116-120.

Want, R. (2006). An Introduction to RFID Technology. *Pervasive Computing*, *5*(1), 25-33. Washington, DC: IEEE Computer Society.

Warf, B. (2006). International competition between satellite and fiber optic carriers: A geographic perspective. *The Professional Geographer, 58*, 1-11.

Warf, B. (2003). Mergers and acquisitions in the telecommunications industry. *Growth and Change, 34*, 321-344.

Warf, B. (1998). Reach out and touch someone: AT&T's global operations in the 1990s. *Professional Geographer, 50*, 255-267.

Watts, C. (2003). http://www/ic.mr.icfai.org/casestudies/catalgoue/Business%20Strategy2/Motorola.ht.

Waverman, et. al (2007). Access Regulation and Infrastructure Investment in the Telecommunications Sector: An Empirical Investigation. London Economics Consulting Group with the Support of ETNO, *mimeo.* September 2007.

Waverman, L., & Dasgupta, K. (2006). Investment in telecommunication networks: A 21st Century perspective. Unpublished manuscript.

WCAG (1999). *Web Content Accessibility Guidelines 1.0.* Retrieved 28th Feb 2008, from http://www.w3.org/TR/1999/WAI-WebCONTENT-19990505/

Webb, W. (2001). *The Future of Wireless Communications.* Boston, MA: Artech House.

Webb, W. (Ed.) (2007). *Wireless Communications: The Future*: Wiley.

Webster, L. (2002). *Motorola University Scores High Grades in China*. Asia Times Online Co. Ltd.

Weeks, M., & Williamson, B. (2006). *A Sound Basis for Evidence Based Policy? A Critique of the ECTA Regulatory Scorecard and SPC Network Papers on Investment and Broadband.* Paper prepared for ETNO by Indepen.

WEFA Group (1995). *Economic impact of deregulating U.S. communications industries.* Bala Cynwyd, PA: The WEFA Group.

Wei, C. P., & Chiu, I. T. (2002). Turning telecommunication call detail to churn prediction: a data mining approach. *Expert System with Application*, *23*, 103-112.

Wei, H., Ganguly, S., & Izmailov, R. (2004). Ad hoc relay network planning for improving cellular data coverage. *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2*, 769-773.

Wei, H.-Y., & Gitlin, R. D. (2005). Incentive scheduling for cooperative relay in WWAN/WLAN two-hop-relay network. *IEEE Wireless Communications and Networking Conference*, *3*, 1696-1701.

Wei, Y., & Bello, D. (2004). *Performance outcomes for business marketers in China: The moderating role of legitimacy behaviors.* ISBM Conference Workshop.

Weiber, R. (1992). *Diffusion von Telekommunikation*. Wiesbaden: Gabler.

Weilenmann, A. (2001). Negociating use: Making sense of mobile technology. *Personal and Ubiquitous Computing, 5*(2), 137-145.

Weiser, M. (1991). The Computer of the 21st Century. *Scientific American*, *265*(3), 66-75.

Weiser, M. (1993). Some Computer Science Issues in Ubiquitous Computing. *Communications of the ACM*, *36*(7), 75-84.

Weiser, P. (2003). Toward a Next Generation Regulatory Strategy. *Loyola University Chicago Law Journal,* 35, 41-85.

Weiss, S. (2002). *Handheld Usability*. Chichester, UK: John Wiley & Sons.

Weiss, T. A. & Jondral, F. K. (2004). Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency. *IEEE*. *Communications Magazine*, *42*(3), S8-S14.

Welch, L., & Benito, G. R. G. (1996). De-internationalization. In *Working papers: International Business Enterprises*. Kingswood, N.S.W.: Department of Marketing, University of Western Sydney.

Wendorf, R. G., & Seidenberg, H. B. (2007). Channel-change games for highly interfering spectrum-agile wireless networks. *2nd International Symposium on Wireless Pervasive Computing*, *1*, 611-616.

Weniger, K., & Zitterbart, M. (2002). IPv6 auto-configuration in large scale mobile Ad-hoc networks, in Proc. of *European Wireless 2002* (pp. 142-148). Florence, Italy, February 2002.

Werbach, K. (2002). A Layered Model for Internet Policy. *Journal of Telecommunications & High-Technology Law, 1*(37), 58-64.

Werbach, K. (2004). Supercommons: Toward a unified theory of wireless communication. *Texas Law Review, 82*(4), 863-973.

Wernerfelt, B. (1984). A Resource-based View of the Firm. *Strategic Management Journal*, *5*(2), 171-180.

Westlaw. *High Plains Wireless, L.P., Appellant v. Federal Communications Commission, Digital PCS, LLC and Tritel Communications, Inc., Intervenors.*

Westlaw. *Omnipoint Communications MB Operations, LLC v. Town of Lincoln D Mass, 2000, 107 F.Supp.2d.*

Weymes, E, (2004). A challenge to traditional management theory, *Foresight, Volume 6* Number 6, 2004, pp. 338-348.

Whalley, J. (2004). Flagship firms, consolidation and changing market structures within the mobile communications market. *Telecommunications Policy, 28*(2), 161-175.

White D. W., Robicheaux R. A., (1995). Linking Strategy Creation Styles With Corporate Climate In A Distribution Channel Context, available at: http://www.sbaer.uca.edu/research/sma/1995/pdf/01.pdf

White, B. (2007). Its Creators Call Internet Outdated Offer Remedies. *The Wall Street Journal Online Edition*, October 2, 2007. Retrieved October 2, 2007.

White, L. J. (2000). 'Propertyzing' the electromagnetic spectrum: why it's important, and how to begin. *Media Law & Policy*, *9*(1), 29-30.

Whitt, R. (2004). A Horizontal Leap Forward: Formulating a New Communications Public Policy Framework Based on the Network Layers Model. *Federal Communications Law Journal, 56*(3), 587-672.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). *An Introduction to Human Factors Engineering*, 2nd edn.. Upper Saddle River, NJ: Pearson Education.

Wickramage, N., & Weerawarana, S. (2005). A Benchmark for Web Service Frameworks. In *Proceedings of the 2005 IEEE international conference on service computing*, (pp. 233-242).

Wietgrefe, H. (2002). Investigation and practical assessment of alarm correlation methods for the use in GSM access networks. *Proceedings of the IEEE/IFIP Network Operations and Management Symposium* (pp. 391- 403). Florence, Italy.

Wietgrefe, H., Tuchs, K., Jobmann, K., Carls, G., Fröhlich P., Nedil, W., & Steinfeld, S. (1997). Using Neural Networks for Alarm Correlation in Cellular Phone Networks. *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications* (pp. 7.2.97.1-10). Melbourne, Australia.

Wild, B., & Ramchandran, K. (2005). Detecting primary receivers for cognitive adio applications. *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 1*, 124-130.

Wiley, M. (1995). LindoApi optimization tool. Site: http://www.lindo.com/

Wilkinson, I., & Young, L. (2000). On cooperating: Firms, relations and networks. *Journal of Business Research, 55*(2), 123-132.

Willers, J. L., Seal, M. R., & Luttrell, R. G. (1999). Remote Sensing, Line-intercept Sampling for Tarnished Plant Bugs (Heteroptera: Miridae) in Mid-south Cotton. *The Journal of Cotton Science, 3*, 160-170.

Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: a review of 40 years of research. *Research in Organizational Behavior, 20*, 77-140.

Williamson, B. (2007, March). *Risk, reward and efficient investment in access networks.* Paper presented at an OFCOM European seminar on regulatory challenges posed by next generation access networks, London, United Kindgom.

Willig, R. (1976). **Consumer's surplus without apology.** *American Economic Review, 66*(4), 589-597.

Wilson, C. (2007). Testing VoIP for quality. *Telephony, 248*(5), 33.

Wilson, C. (2005). VoIP quality questions stir echoes. *Telephony, 246*(12), 10.

Wilson, J. R. (2000). Fundamentals of Ergonomics in theory and practice. *Applied Ergonomics, 31*(6), 557-567.

Wilson, J. R., & Corlett, N. (Eds.), (2005). *Evaluation of Human Work*, 3rd edn. London, UK: Taylor & Francis.

WiMax-Forum. (2004). *Business Case Models for Fixed Broadband Wireless Access based on WiMAX Technology and the 802.16 Standard.* Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads

WiMax-Forum. (2004). *Regulatory position and goals of the WiMAX forum*. Retrieved May 14, 2008, from http://www.wimaxforum.org/news/downloads/WiMAX_Forum_Regulatory_Whitepaper_v06162004.pdf.

WiMAX-Forum. (2006). Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation.

WiMax-Forum. (2007). *IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access.* Retrieved May 14, 2008, from http://www.wimaxforum.org/news/downloads/WiMAXWhitepaper.pdf.

WiMax-Forum. (2007a). *egulatory Position and Goals of the WiMAX Forum*. Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads/WiMAX_Forum_Regulatory_Whitepaper_v06162004.pdf.

WiMax-Forum. (2007b). *IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access. available at.* Retrieved, from the World Wide Web: http://www.wimaxforum.org/news/downloads/WiMAXWhitepaper.pdf.

Wireless Asia (2002). Virgin Mobile and SingTel jointly call off their Virgin-branded MVNO mobile service a mere eight months after its launch. August 21 2002.

Wirth, M. (2006). Issues in media convergence. In A. Albarran, S. Chan-Olmsted, & M. Wirth (Eds.), *Handbook of media management and economics* (pp. 445-462). Mahwah, NJ: Lawrence Erlbaum & Associates.

Wirtz, B. W. (2001). Reconfiguration of Value Chains in Converging Media and Communication Market. *Long Range Planning, 34*, 489-506.

Wong, C. C., & Hiew, P. L. (2005). Correlations between factors affecting the diffusion of mobile entertainment in Malaysia. *ICEC 2005*, 615-621.

Wong, K. Y. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial Management & Data Systems, 105*(3), 261-279.

Wong, K. Y., & Aspinwall, E. (2003). *Is knowledge management equivalent to information technology?* Paper presented at the Fourth European Conference on Knowledge Management (989-997), Oxford University, Oxford.

Wong, V. W. S., & Leung, V. C. M. (2000). Location Management for Next-Generation Personal Communications Networks. IEEE Network, (pp.18-24).

Wood, J., & Silver, D. (1995) *Joint Application Development*, 2nd edn. New York, NY: John Wiley & Sons.

Woolfson, M. (2005). Can self-regulation protect the mobile industry from crooks? *New Media Age, 17/03/2005*, 16.

World Trade Organization (1998), Telecommunication Services, Background Note by the Secretariat, S/C/W/74, 8 Dec.http://203.127.220.68/apecp1.nsf/websearch/D5038CDD0095E0348256BF400160256?OpenDocument

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557-585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161-215.

Wright, S. (1960). Path coefficients and path regression: Alternative or complementary concepts? *Biometrics, 16*, 189-202.

Wu, B., & Wang, Q. (1995). A multiobjective analytic framework for slotted ALOHA wireless LANs. *4th IEEE International Conference on Universal Personal Communications, 1*, 12-16.

Wu, F. (2003). From monopoly to competition: a comparative study of telecommunications policy reform in four Asia countries. *Proceeding of Pacific Telecommunications Council*, January 2003, Manoa,Hawaii.

Wu, J.-H., & Wang, S.-C. (2005). What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. *Information and Management, 42*(5), 719-729.

Wu, T., & Somani, A. (2006). Attack monitoring and localization in all-optical networks. *Cluster Comput, 9*, 465-473.

Wu, T., & Somani, A. (2005). Cross-talk attack monitoring and localization in all-optical networks. *IEEE/ACM Transactions on Networking, 13*(6), 1390-1401.

Wu, Y., & Pierre, S. (2003). Optimization of Access Network Design in 3G Networks. *Canadian conference on electrical and computer engineering (CCECE)* (pp. 781-784).

Xenikou A. & Simosi M., (2006). Organizational culture and transformational leadership as predictors of business unit performance, *Journal of Managerial Psychology* 21, 6, pp. 566-579

Xie, H., Krishnamurthy, A., Silberschatz, A. and Yang, Y. P4P: Explicit communications for cooperative control between P2P and network providers, <www.dcia.info/documents/P4P_Overview.pdf>

Xiong, K., & Perros, H. (2006a). Security-aware resource allocation in Web services. In *Proceedings of the International Conference on Web Services*, Chicago, IL.

Xiong, K., & Perros, H. (2006b). Resource optimization subject to a percentile response time SLA for enterprise computing. In *Proceedings of the IEEE Globecom: Quality Reliability and Performance Modeling for Emerging Network Services (Globecom-QRPM),* San Francisco, CA.

Xiong, K., & Perros, H. (2008). SLA-based service composition in enterprise computing. In *Proceedings of International Workshop on Quality of Service*. Enschede, The Netherlands.

Xun, Y., Chee, K., Siew, X., Feng, W., & Eiji, O. (2000). A Secure Agent-based Framework for Internet Trading in Mobile Computing Environments. *Distributed and Parallel Databases, 8,* 85-117. MA, USA: Kluwer Academic Publishers Hingham.

Yan, L., Dodier, R., & Mozer, M. (2003). Optimizing Classifier Performance via an Approximation to the

Wilcoxon-Mann-Whitney Statistic. In F. Tom & M. Nina (Eds.), *Proceeding of the Twentieth international Conference on Machine Learning*, (pp. 848-855). Washington: *AAAI Press.*

Yan, L., Wolniewicz, R., & Dodier, R. (2004). Predicting customer behaviour in telecommunications. *IEEE Intelligent Systems*, *19*, 50–58.

Yang, D.-N., & Liao, W. (2003). Design of light-tree based logical topologies for multicast streams in wavelength routed optical networks. *Proceedings of IEEE INFOCOM*.

Ye, Y., Chai, T., Cheng, T., & Lu, C. (2003). Agorithms for the design of WDM translucent optical networks. *Optics Express, 11*(22), 2917-2926.

Yemini, S. A., Kliger, S., Mozes, E., Yemini, Y., & Ohsie, D. (1996). High speed and robust event correlation. *IEEE Communications Magazine*, 34, 82-90.

Yi, Y. H., & Lee, C. L. (2005) Antecedents and consequences of switching costs: The moderating role of service subscription Types. *Journal of Korean Marketing Association*, *20*(3), 1-28.

Yin, R. K. (1994). *Case Study Research: Design and Methods.* Beverley Hills: Sage.

Ying, H., Ching, W., Wang, L., Hua, H., Wen, H., & Lo, H. (2003). Applying Mobile Agents to E-business. *Tamkang Journal of Science and Engineering*, *6*(3), 159-172.

Yingying, T., Wade, M., & Richard, P. (2007). Detecting and localizing wireless spoofing attacks. In R. Cruz (Ed.), *The Fourth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks* (pp. 193-202). IEEE: USA.

Yip, G. S. (1989). Global Strategy…In a World of Nations? *Sloan Management Review, 31*(1), 29-41.

Yip, G. S. (2000). Global Strategy in the Internet Era. *Business Strategy Review, 11*(4), 1-14.

Yoffie, D.B., & Bergenstein, S. (1985). Corporate political entrepreneurship. *California Management Review, 28*, 1, 313-337.

Yoo, C. (2005). Beyond Network Neutrality. *Harvard Journal of Law & Technology, 19*(1), 1-77.

Yoo, C. S. (2006). Network neutrality and the economics of congestion. *Georgetown Law Journal, 94*(6), 1847.

Yoo, Y., Lyytinen, K., & Yang, H. (2005). The role of standards in innovation and diffusion of broadband mobile services : The case of South Korea. *Journal of Strategic Information Systems*, *14*(3), 323-353.

Youngblood, M. D. (2000). Winning cultures for the new economy, *Strategy and Leadership, 28*(6), 4-9.

Yu, O., & Cao, Y. (2006). Placement of light splitters and wavelength converters for efficient multicast in all-optical WDM networks. *IEICE Trans. Inf. & Syst., E89-D*(2), 709-717.

Yun, Y. B., Nakayama, H., Tanion, T., & Arakawa, M. (2001). Generation of efficient frontiers in multi-objective optimization problems by generalized data envelopment analysis. *European Journal of Operational Research*, *129*, 586-595.

Yusof, A. M. (1998). *The future path of Malaysian telecommunication industry.* Retrieved March 18, 2005, from http://members.tripod.com/~azman97/essay7.html

Zeltserman, D. (1999). *A Practical Guide to SNMPv3 and Network Management.* New Jersey: Prentice Hall.

Zeng, P-Y. (2007). http://news.xinhuanet.com/english/2007-09/12/content_6711568.htm

Zhang, A., Melcher, A., & Li, L. (2004). Mapping the relationships among product complexity, information technology, and transaction governance structure: A case study. *Journal of Management Systems*, *16*(4), 41-55.

Zhang, D., & Tian, Z. (2007). Adaptive games for agile spectrum access based on extended Kalman filtering. *IEEE Journal of Selected Topics in Signal Processing, 1*(1), 79-90.

Zhang, L., Liang, Y.-C., & Xin, Y. (2007). Joint admission control and power allocation for cognitive radio networks. *IEEE International Conference on Acoustics, Speech and Signal Processing, 3*, 673-676.

Zhang, M., & Wolff, R. S. (2004). Crossing the digital divide: cost-effective broadband wireless access for rural and remote areas. *IEEE Communications Magazine, 42*(2), 99- 105.

Zhang, Q., Yu, T., & Irwin, K. (2004). A classification scheme for trust functions in reputation-based trust management. *In Proceedings of the International Workshop on Trust, Security, and Reputation on the Semantic Web*, Hiroshima, Japan.

Zhao, B., Huang, L., Stribling, J., Rhea, S., Joseph, A., & Kubiatowizc, J. (2003). Tapestry: A resilient global scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, *22*(1), 41-53.

Zheng, S., Yen, D., & Tarn, J.M. (2000). The new spectrum of the cross-enterprise solution: The integration of supply chain management and enterprise resource planning systems. *The Journal of Computer Information Systems, 41*(2), 84-93.

Zhong, L., & Ramamurthy, B. (2001). Optimization of amplifier placements in switch-based optical networks. *IEEE International Conference on Communication*, 224-228.

Zhou, H., Mutka, M., & Ni, L.M. (2004). IP address handoff in MANET. *Computer Communications, 23*(1), 2425-2433.

Zhou, H., Ni, L. M., & Mutka, M. W. (2003). Prophet address allocation for large scale MANETs. *Ad Hoc Networks Journal*, *1*(4), 423-434.

Ziegler, C., & Lausen, G. (2004). Spreading Activation Models for Trust Propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-commerce, and e-Service (EEE '04).*

Ziv, N. D., & Mulloth, B. (2006). An Exploration on Mobile Social Networking: Dodgeball as a Case in Point. *International Conference on Mobile Business.* Washington, DC: IEEE Computer Society.

About the Contributors

Angeliki Z. Agorogianni was born in Larisa, Greece on May 8, 1973. She received the Dipl. degree in electrical engineering from the Aristotle University of Thessaloniki, Greece and the Master in Business Administration from the University of Macedonia, Greece in 1996 and 2001 respectively. Since 1997 she has been working in the Telecommunications Center of the Aristotle University of Thessaloniki, Greece. Her research interests include distance-learning technologies, data modeling, service oriented architectures and enterprise architecture planning.

Cajetan M. Akujuobi received the BS degree from Southern University, Baton Rouge, LA, in 1980, theMS degree from Tuskegee University, AL, in 1983, and the PhD degree from George Mason University, Fairfax,VA, in 1995, all in electrical engineering, and the MBA degree from Hampton University, Hampton, VA, in 1987. He is a professor in the Department of Electrical Engineering and is the founding director of Analog and Mixed Signal, DSP Solutions and High Speed (Broadband) Communication Programs at Prairie View A & M University, Prairie View, TX. He is also the founding director of the Center of Excellence for Communication System Technology Research. His research interests include signal/image processing and communication systems (broadband telecommunications) using such tools as wavelet and fractal transforms; compressed sensing/compressive sensing. His other research interests are in the areas of digital signal processor solutions, analog mixed-signal systems, and control system-based communications. He was a Participant and collaborative Member of ANSI TIE1.4 Working Group that had the technical responsibility of developing T1.413, Issue 2 ADSL standard. He has published extensively and has also written many technical reports. He was selected as one of the U.S. representatives for engineering's educational and consultation mission to Asia in 1989. Prof. Akujuobi is a Senior Member of ISA, ASEE, SPIE, and Sigma Xi, the Scientific Research Society. He is one of the founding corporate members of the IEEE Standards Association (IEEE-SA), Industry Advisory Committee (IAC).

Nana K. Ampah is currently a post-graduate student at Prairie View A & M University and also an Electrical Engineer with Jacobs Engineering, Houston, Texas. He is a student member of IEEE and belongs to organizations such as Ghana Institute of Engineers - GhIE (Associate Member) and Ghana Institute of Management and Public Administration (GIMPA) Alumni Association (Member). Nana K. Ampah graduated from Prairie View A & M University with anMSE.E. in Communications in May, 2004. He also graduated from Kiev Polytechnic Institute (KPI) in Kiev, Ukraine in June, 1993 with anMSE.E. specializing in Power Systems and Networks. He was also awarded a Post-graduate Certificate in Urban Management by GIMPA in June, 2001. He worked with the Electricity Company of Ghana for 8 years and has over 6 years experience in the design, construction, commissioning and management of urban and rural electrification projects. He was a project engineer under a World Bank funded urban/rural electrification project from 1996 to 2002. He also worked with Skanska Jensen International as a Materials Coordinator (Consultant) on a World Bank funded rural electrification project in 1997. He is currently doing research work in the communications network security area involving the development of Intrusion Detection and Prevention Systems (IDS/IPS) for Homeland Security and Enterprise Networks.

Georgios N. Angelou was educated at Democritus University of Thrace, Greece and received the Diploma degree in Electrical Engineering, in 1995. He also received a MSc in communications & radio Engineering from King's College London in 1997 and a MSc in techno-economics from the National Technical University of Athens in 2002. He worked in telecommunication industry, in companies Siemens and Nokia, for four years as a technical consultant. He is a PhD candidate in information systems at the University of Macedonia, Thessaloniki. His current research interests are ICT business analysis and technology investments evaluation.

Nick Antonopoulos, Dr. Nick Antonopoulos is currently a senior lecturer (US Associate Professor) at the Department of Computing, University of Surrey, UK. He holds a BSc in physics (1st class) from the University of Athens in1993, an MSc in information technology from Aston University in 1994 and a PhD in computer science from the University of Surrey in 2000. He has over 9 years of academic experience during which he has designed and has been managing advanced Master's Programmes in computer science at the University of Surrey. He has published over 60 research articles in distributed systems, peer-to-peer networks and software agents.

Luigi Atzori (l.atzori@diee.unica.it) is assistant professor at the University of Cagliari (Italy) since 2000. His main research topics of interest are in multimedia networking: error recovery and concealment, IP Telephony, video streaming, network QoS management. He has published more than 60 journal articles and refereed conference papers. Dr. Atzori has been awarded a Fulbright Scholarship (11/2003-05/2004) to work on video at the University of Arizona. He is editor for the ACM/Springer Wireless Networks Journal and is involved in the organization of several International Conferences on Multimedia Networking.

Raquel Barco received her MSc degree in telecommunication engineering in 1997 and her PhD in 2007 from the University of Málaga (Spain). From 1998 to 2000 she worked at the European Space Agency in Darmstadt (Germany). From 2000 to 2003 she worked part-time for Nokia Networks. Since the end of 1999 she has been with the Communication Engineering Department, University of Málaga. Her research interests include satellite and mobile communications, mainly focused on self-management of radio access networks.

Georges Baume, Dr. Georges Baume is a senior lecturer in the school of business at the University of Adelaide. He lectures in strategic management, corporate strategy, and corporate transformation, and is the strategy discipline leader for the school. Dr. Baume's research interests include the development, implementation, and execution of internationalisation strategies and processes for service sector firms.

He holds a PhD and an MBA from the University of Adelaide, and an M.Sc from Oxford University. He has won numerous teaching excellence awards for his work in MBA and executive education programs in Australia, Asia, and Europe.

Lokesh B. Bhajantri received M.Tech degree in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belgaum, Karnataka, India, in 2005. He is presently working as a lecturer in the Department of Information Science and Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India. His areas of interest include m-commerce, e-commerce, networking protocols, applications of mobile agents and real time systems. He has published few papers in national/international conferences.

David Bell holds a BS degree in electrical engineering from Walla Walla College and aMS degree in Business Administration with an emphasis in Management Information Systems from California State University, Sacramento. He is currently pursuing a PhD in applied management and decision sciences with an emphasis in Information Systems Management from Walden University. For the past 18 years he has been teaching a wide variety of information systems courses at Pacific Union College, a private college in California's Napa Valley. He enjoys exploring new technologies, photography, singing, and scuba diving.

Fabrizio Boi (f.boi@diee.unica.it) received his Dr. Eng. Degree in electronic engineering from the University of Cagliari, Italy, in 2006. In 2006 he started his PhD research activities with the Department of Engineering (DIEE) at the University of Cagliari. In 2007 he taked part of IKNOS research project as collaborator of DIEE which partners were Tiscali Italia SpA, CNIT and DIEE. In 2007 he joined in Open Media Center Lab of Sardegna DistrICT as collaborator at Sardegna Ricerche, Italy. His research interests are related to IP Telephony playout buffering, speech quality estimation (QoS on IP networks), NGNs performance management, home entertainment, home automation and multi-touch sensing devices.

Mohamed Boucadair was an R&D engineer with France Telecom R&D in charge of dynamic provisioning, QoS, multicast, and intra/interdomain traffic engineering, and now works on VoIP services. He graduated from the Ecole Nationale Supérieure d'Ingénieur de Caen (Institut des Sciences de la Matière et des Rayonnements), a French school of engineers.

Susan J. Chinburg received a BS Science Teaching and Mathematics from Minnesota State University, Mankato, MN in 1972, aMS in geology from San Jose State University, San Jose, CA in 1985 and a PhD in business administration in MIS and telecommunications in 2006. Currently she is an assistant professor at Rogers State University, Claremore, OK. Her research interests include telecommunications network analysis and design, applied applications of MIS techniques to other areas and information assurance and security applications and teaching in the online environment. She has worked for Conoco, Inc in exploration, development, research and systems analysis and for WilTel, (formerly Williams Communications Group) in optical network architecture.

Chin Wei Chong is currently a lecturer in Multimedia University, Malaysia. She received her PhD in the area of knowledge management. She has published several papers in international refereed journals, conference proceedings and book chapters. Her research interests include knowledge management, human resource management, strategic management and multimedia learning system.

Siong Choy Chong received his PhD from Multimedia University. He has published extensively in international refereed journals, conference proceedings and book chapters in addition to serving on a number of journal editorial boards. His research interests include knowledge management, human resource management, information technology management and entrepreneurship. He is currently the chief executive officer of Putra International College, Malaysia.

Nicoletta Corrocher, Laurea in economics (Bocconi University, Milan), MSc in science and technology policy (SPRU, Science and Technology Policy Research, University of Sussex, Brighton, UK), PhD in economics and management of innovations (Sant'Anna School of Advanced Studies, Pisa), is lecturer at Bocconi University and Research Fellow at CESPRI (Research Centre on Processes of Innovation and Internationalisation). Her main research interests include technology diffusion, product and service innovation and industrial dynamics in the ICT sector, with particular reference to the telecommunications industry.

Rebecca De Coster, Dr. Rebecca De Coster is a lecturer of innovation in the School of Engineering and Design, Brunel University. She has twelve years experience in the telecoms sector including research and consultancy based on her prior experience in telecommunications planning. She received her BEng (Hons) in electronic and electrical engineering from Birmingham University and her MBA from the University of Kingston, London. She earned her doctoral degree from Brunel University examining innovation strategies of mobile networking firms and the associated enterprise realignment for managing technology and innovation.

Walter Distaso is at Tanaka Business School since September 2006. He holds a PhD from the University of York. Previously, he held positions at the University of Exeter and Queen Mary, University of London. He has also been a visiting professor at the IMF. His research interests are in the area of estimation, specification testing and prediction of financial Volatility in continuous time models; analyzing macroeconomic and financial time series using long memory models; identifying the macroeconomic determinants of stock-market volatility; studying the dependence of multivariate financial time series using copulas; evaluating competing trading strategies; analyzing the features and the effects of market microstructure noise; analysis of telecommunications market. His research has been published by leading journals such as *Review of Economic Studies*, the *Journal of Econometrics* and the *Journal of Business and Economic Statistics.*

Anastasios A. Economides received the Dipl.Eng. degree in electrical engineering from the Aristotle University of Thessaloniki, in 1984. Holding a Fulbright and a Greek State Fellowship, he received the MSc and the PhD degrees in computer engineering from the University of Southern California, Los Angeles, in 1987 and 1990, respectively. At graduation, he received the Outstanding Academic Achievement Award from the University of Southern California. Currently, he is an associate professor of computer networks at the University of Macedonia, Thessaloniki, Greece. He is the director of CONTA (COmputer Networks and Telematics Applications) Laboratory. His research interests include networking technologies and e-services. He has published over one hundred peer-reviewed papers. He has been the plenary speaker in two International Conferences. He has served on the program committees for many International Conferences, and as a reviewer for many International Journals and Conferences. Finally, he has been the principal investigator of several funded projects and participated in many funded projects.

Salah Eddine Elayoubi was born in Tripoli, Lebanon, in 1978. He received the engineering diploma in telecommunications and computer science from the Lebanese University in 2000 and the master's degree in telecommunications and networking from the National Polytechnic Institute at Toulouse, France, in 2001. He completed his PhD degree in computer science and telecommunications in the University of Paris VI in 2004. Since then, he has been working in Orange Labs. His research interests include radio resourse management and performance evaluation of mobile networks.

Jose Joaquin Escudero-Garzas received the telecommunication engineer degree from Polytechnic University of Madrid (Spain) in 1997. He worked for different telecommunications companies until 2002. He joined University Carlos III de Madrid in 2002, where he is currently working as an assistant lecturer and working towards the PhD in electrical engineering. He has participated in several national and international research projects. His research areas of interest include resources management for wireless communications, cognitive radios and game theory for wireless networks.

Károly Farkas received his PhD degree in computer science in 2007 from ETH Zurich, Switzerland, and his MSc degree in Computer Science in 1998 from the Budapest University of Technology and Economics (BME), Hungary. Currently he is working as an associate professor at University of West Hungary, Sopron, Hungary. His research interests cover the field of communication networks, especially autonomic, self-organized and wireless mobile ad hoc networks. His core research area deals with service provisioning in mobile ad hoc networks. He has published more than 30 scientific papers in different journals, conferences and workshops and he has given a plenty of regular and invited talks. In the years past, he supervised a number of student theses, participated in several research projects, coordinated the preparation of an EU IST research project proposal and acted as reviewer and organizer of numerous scientific conferences. He served in the program committee of the IADAT-tcn conference in 2005 and 2006, and he acted as a technical co-chair of the 3rd WICON (Wireless Internet) conference in 2007. He is a member of IEEE and fellow of the European Multimedia Academy.

Claudio Feijóo holds a MSc and PhD in telecommunication engineering from Universidad Politécnica de Madrid, Spain. Currently, he is working at the Institute for Prospective Technological Studies of the JRC of the European Commission where he researches on the future socio-economic impact of emerging information society technologies. He is professor (on leave) in the Universidad Politécnica de Madrid, where he directed the Chair in Telecommunications Regulation and Information Society Public Policies. He was also adviser for the Spanish State Secretary on Telecommunications and Information Society, participating in the design of information society development plans. He lectures regularly in seminars and postgraduate courses and has authored many publications in journals and conferences.

Zulima Fernández is professor of strategic management and holds the Chair of Entrepreneuship and Family Firms Management at University Carlos III de Madrid (Spain). She received her PhD from University of Oviedo (Spain). Her research focuses in strategic management, technology management,

management and privatization of state-owned companies and organization design. Recent papers deal with strategic decisions of SMEs, privatization and first-mover advantages. An example of her empirical research can be found in journals such as *Journal of International Business Studies*, *Journal of International Entrepreneurship*, among others

Tiago Fioreze is a PhD student at the University of Twente, Enschede, the Netherlands. He is currently investigating the use of self-management in optical networks. He graduated in Computer Science from the Federal University of Rio Grande do Sul (UFRGS) in 2005 (Master of Science) and from the Federal University of Santa Maria (UFSM) in 2003 (Bachelor of Science). His main topic of interest is Network Management, but he also has interest in areas such as Network Security and Fault Tolerance.

Roberto Fontana is currently assistant professor in applied economics at the University of Pavia and Research Fellow at CESPRI – Bocconi University. He received his PhD in science and technology policy studies at SPRU – University of Sussex, UK in 2002 and his master of science in science and technology policy studies from SPRU – University of Sussex in 1998. His main interests are in the field of innovation diffusion and standardisation, product innovation, variety and its implications for the dynamics and structural evolution of industries with particular reference to high technology sectors such as computers and communication equipment.

Ana Garcia-Armada received the telecommunication engineer degree from the Polytechnic University of Madrid (Spain) in July 1994 and PhD in electrical engineering from the same, February 1998. She is currently working as an associate professor at the University Carlos III de Madrid, Spain where she has occupied several management positions. She has participated in several national and international research projects. She is co-author of four books on wireless communications and signal processing. She has published 22 papers in international journals and more than 40 contributions to international conferences. She has contributed to international organizations such as ITU and ETSI.

Richard A. Gershon, PhD, (Ohio University) is professor and co-founder of the Telecommunications & Information Management program at Western Michigan University where he teaches courses in telecommunications management and communication technology & innovation. Dr. Gershon is the author of *Telecommunications & Business Strategy* (2008 - in press) and *The Transnational Media Corporation: Global Messages and Free Market Competition*, winner of the 1998 book of the year sponsored by the National Cable Television Center. Dr. Gershon has twice been selected for teaching honors at the national level, including the Steven H. Coltrin Professor of the Year Award (2000) by the International Radio& Television Society (IRTS) and the Barry Sherman Award for Teaching Excellence (2001) by the Management and Economics division of the Association for Education in Journalism and Mass Communication (AEJMC). In 2007, he was the recipient of the Distinguished Teaching Award at Western Michigan University. Dr. Gershon is a founding member of the International Telecommunications Education & Research Association (ITERA) and served as the organization's first President.

David Golightly, holds a doctorate in psychology from the University of Nottingham in ICT-based problem solving. He also has a BSc in psychology (Aberdeen), an MSc in Cognitive Science (Cardiff), and is a chartered psychologist. He has substantial commercial experience in developing and managing human factors research programmes, having lead human factors and systems usability activities in both the telecoms and financial services sector. David joined the Human Factors Research Group in 2007, where he continues to manage research programmes exploring human factors in domains such as web, mobile, transport and requirements management.

José Luis Gómez-Barroso, holds a PhD and a MSc in economics from Universidad Nacional de Educación a Distancia (UNED). He also holds a MSc in telecommunication engineering from Universidad Politécnica de Madrid as well as another MSc in Law from Universidad Complutense. Dr. Gómez-Barrroso is an associate professor in the Department of Applied Economics and Economic History at Universidad Nacional de Educación a Distancia (UNED). His research interests are focused on telecommunications regulation and public aspects of the development of Information Society. Dr. Gómez-Barroso has participated in different research projects, some of them for the European Commission. He is the co-ordinator of the Spanish Group for the Study of Regulation of Telecommunications (GRETEL).

Ricardo Gonçalves is an assistant professor at the Faculty of Economics and Management of the Portuguese Catholic University, in Porto. He has done both consultancy work and academic research on various regulatory and competition aspects of the telecommunications sector. He previously worked in a London based economic consultancy firm. He holds a PhD and a MSc in Economics from the University of York (UK) and graduated in Economics from Instituto Superior de Economia e Gestão (UTL, Portugal). His main areas of expertise are auctions, microeconomics, industrial organisation, and competition policy and regulation.

Sotirios K. Goudos was born in Thessaloniki, Greece in 1968. He received the BSc degree in physics in 1991 and the MSc degree in Electronics in 1994 both from the Aristotle University of Thessaloniki. In 2001 he received his PhD degree in physics from the Aristotle University of Thessaloniki and in 2005 the master's in information systems from the University of Macedonia, Greece. Since 1996 he has been working in the Telecommunications Center of the Aristotle University of Thessaloniki, Greece. His research interests include antenna and microwave structures design, electromagnetic compatibility of communication systems, evolutionary computation algorithms and semantic web technologies.

Lisandro Zambenedetti Granville is an associate professor at the Institute of Informatics of the Federal University of Rio Grande do Sul (UFRGS), Brazil. He received his MSc and PhD degrees, both in computer science, from UFRGS in 1998 and 2001, respectively. He has served as a TPC member (2003-2008), General Co-Chair (2004), and Steering Committee member (2005-2008) for the Brazilian Symposium on Computer Networks (SBC/LARC SBRC). Currently, he is member of the Brazilian Internet Committee (CBG.br). He has served as a TPC member for IFIP/IEEE IM 2005, NOMS 2006, IM 2007, NOMS 2008 and IFIP/IEEE DSOM 2005, DSOM 2006, DSOM 2007 (TPC Co-Chair), DSOM 2008. His main areas of interest include policy based network management, management using/of Web services, network monitoring and configuration, and P2P-based services and applications.

Hyun-Soo Han is an associate professor of information technology management at College of Information and Communications, Hanyang University in Seoul, Korea. He received a BS in industrial engineering from Seoul National University, andMS in management science at Korea Advanced Institute of Science and Technologies (KAIST). He earned PhD in management from University of

Massachusetts Amherst, USA. Before joined as a faculty member of Hanyang University, he had been with a Korean IT service company as a director of the IS consulting service department. His technical publications appear in various international and domestic journals such as *Information & Management*, *International Journal of Satellite Communications and Networking, International Journal of Technology Management*, *International Journal of Operations and Quantitative Management*, *European Journal of Operational Research*, *Annals of Operations Research*, and others. His recent research interests include industry convergence, mobile applications, e-commerce, information systems and telecommunication management.

Franz Hartleb received the Diplma degree in computer science from the University of Erlangen-Nürnberg (Germany) in 1988 and the Doctor of Engineering degree in 1992. From 1993 to 1995 he was affiliated with Siemens AG. Since 1996 he is a senior expert in traffic management and network optimization at the research and development center of Deutsche Telekom (now T-Systems). Currently he is working on IP traffic analysis, traffic modeling, link dimensioning, and QoS guarantee.

Gerhard Haßlinger has more than 10 years of experience as a researcher at the computer science department of Darmstadt University of Technology in Germany, as well as in telecommunications industry at T-Systems, where he is currently involved in the design and planning of the broadband access platform of Deutsche Telekom for delivering Internet services with QoS support. His research interests include performance and quality of service aspects in routing, traffic engineering, and peer-to-peer networking.

Sunil Hazari is associate professor in the Department of Management, Richards College of Business, University of West Georgia. His teaching and research interests are in the areas of business education, information security, infrastructure design of e-Commerce sites, Web usability, and organizational aspects of e-learning. He has authored several peer-reviewed journal publications in information and instructional technology areas, has presented papers at national conferences, and is editorial board member of several information system journals. Web site: http://www.sunilhazari.com/education

Daniel Heil holds a master's in public policy candidate in the School of Public Policy at Pepperdine University. He is specializing in economics and American government. His research interests include telecommunications regulations and the effects of federal programs on poverty and employment. He is currently a research intern with the Hoover Institution. He holds a BA in economics from California State University, Northridge.

Trevor Hopper is professor of management accounting at Manchester Business School, and visiting professor at Victoria University of Wellington and Stockholm School of Economics, having previously worked at Wolverhampton and Sheffield Universities, as a cost accountant in industry, and as a visitor at the University of Michigan, Ann Arbor, U.S.A; Queen's University, Canada; Griffith University, Australia; and Kyushu and Fukuoka Universities, Japan. He has co-edited four books and the British Accounting Review, and is on the editorial boards of several major research journals. He has published in professional journals, books, and research journals including *Accounting and Business Research; Accounting, Organizations and Society; the Auditing, Accounting and Accountability Journal; British Accounting Review;*

Critical Perspectives on Accounting; European Accounting Review; Journal of Management Studies; and Management Accounting Research

Zoltán Horváth received his MSc degree from the Budapest University of Technology and Economics (BME), where he actually pursues his PhD studies and gives computer networks courses. He has participated in many R&D projects related to planning, testing and building local and metropolitan area networks including WiMAX and community network technologies, as well. He also worked as an advisor for the Hungarian National Communications Authority in applying ETSI regulations, testing devices, EMC measurements and interference examination cooperating with the Hungarian Meteorological Service.

Stefan Hüsig has a PhD in business administration and holds the position of assistant professor at the Chair of Innovation and Technology Management at the University of Regensburg, Germany. In addition, he works as a consultant in the area of innovation management, where he has worked with Vodafone R&D, deep innovation, Mannesman Pilotentwicklung, Fraunhofer Technology Development Group and other firms. His research interests are the areas of radical/disruptive innovations, telecommunications, predevelopment, corporate entrepreneurship, and computer-aided innovation.

Wesley J. Johnston is the CBIM RoundTable professor of marketing and director of the Center for Business and Industrial Marketing in the Robinson College of Business at the Georgia State University. His research has been published in the *Journal of Marketing*, the *Journal of Consumer Research*, the *Journal of International Business Studies*, the *Journal of Business-to-Business Marketing* and numerous other publications. He was selected as one of the best researchers in marketing by a recent poll of marketing department chairman in U.S. universities. His number of publications places him in the top three researchers in business-to-business marketing in the world.

Mahabaleshwar S. Kakkasageri completed his MTech from Visvesvaraya Technological University Belgaum, Karnataka. He is pursuing his PhD in the area of Vehicular Ad hoc Networks (VANETs). Presently he is serving as an assistant professor in Department of Electronics and Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka. He has published several papers in conferences and journals. Recently he has been listed in Marqui's *Who's Who in the World*.

Faouzi Kamoun assistant professor and assistant dean in the College of Information Technology at the University of Dubai, holds a PhD degree in electrical & computer engineering from Concordia University and an MBA from McGill University. Prior to joining the University of Dubai in 2002, he has been a senior technical advisor with Nortel Networks where he made various contributions in the specifications and rollout of many optical networks. Dr. Kamoun current research is focused on the management aspects of communications networks and technologies. Dr. Kamoun was the recipient of Nortel Networks CEO top talent award in 2000 and 2001. He also received the best faculty award from the University of Dubai in 2004 (a university wide competition). He is a member of the Editorial board of the I*nformation Technology Journal* and *The International Arab Journal of e-Technology (IAJeT).*

Dimitris N. Kanellopoulos holds a PhD degree in multimedia communications from the Department of Electrical and computer engineering of the University of Patras, Greece. He is a member of the Educational Software Development Laboratory (ESDLab) in the Department of Mathematics at the University of Patras. His research interests include multimedia communications, intelligent information systems, knowledge representation and web-based education. He has authored numerous papers in international journals and conferences in these areas. He serves as an editor in ten scientific journals.

Georgia M. Kapitsaki is a research associate working towards a PhD at the National Technical University of Athens (NTUA). Her research interests include mobile internet, context-aware services, context modelling and management, software engineering approaches for pervasive computing and service-oriented architectures. She received her Dipl.-Ing from the School of Electrical and Computer Engineering of NTUA in 2005.

Antonios D. Kargas received his BS degree in economics and his MSc degree in economics and management of telecommunications at the National and Kapodistrian University of Athens. Currently he is working towards his PhD on the organizational issues of telecommunications industry. His research interests include organizational culture and performance analysis of telecommunications companies as well as economics of telecommunications.

Sebastian Kempken has completed his studies in applied computer science at the university of Duisburg-Essen with a thesis on the evaluation of spelling variants by string distance measures. Currently, he is working on his PhD in the project "Verdikom" on the subject of modeling and interval-based analysis of correlated data traffic in service-integrating communication networks.

Moon-Koo Kim is a senior researcher in Technology Strategy Research Division of Electronics and Telecommunications Research Institute (ETRI). He received the BA degree in business administration from Yonsei University, Seoul, Korea and the MS degree in management from Information and Communications University (ICU), Daejeon, Korea. He is mainly interested in marketing strategies, forecasting and business modeling, competitiveness, and IT policy.

Moon-Soo Kim is an associate professor of School of Industrial & Management Engineering, Hankuk University of Foreign Studies (HUFS) in Korea. He has experienced as a project manager at ETRI (Electronics and Telecommunications Research Institute) for 5 years before joining the university. His research focuses on technology strategy and its various application fields. He has been involved in studies on management of technology, telecommunication service management, and e-business planning & strategy, etc. Dr. Kim has published papers in several international journals such as *International Journal of Innovation Management*, *Scientometrics*, *Technology Analysis & Strategic Management*, *Omega*, *Journal of Scientific & Industrial Research*, and *Technological Forecasting & Social Change*, etc., and also in several domestic journals. He holds a PhD from Seoul National University in Korea.

Tobias Kollmann is professor for e-business and e-entrepreneurship at the University of Duisburg-Essen, Germany. Prior to his current position, he was a founder of autoscout24, the largest used car electronic trading platform in Europe, and held a professorship for e-business at the University of Kiel, Germany. Professor Kollmann received his doctorate in 1997. Since then, his research interests include e-business, e-commerce, virtual marketplaces and, in particular, business venturing in the Internet economy.

Stan Kurkovsky is an associate professor at the Department of Computer Science at Central Connecticut State University. Stan earned his PhD from the Center for Advanced Computer Studies of the University of Louisiana in 1999. The results of his doctoral research have been applied to network planning and industrial simulation. Stan's research interests are in mobile commerce and pervasive computing, distributed systems and software engineering. He published over 40 papers in refereed proceedings of national and international conferences, scientific journals and books. Stan serves as a reviewer and a member of program committees on a number of national and international conferences. During his academic career, Stan received over a million dollars in funding from private and federal sources.

Riku Laanti has MSc (international business) and MBA (finance) degrees from the Helsinki School of Economics, and he is currently finalising his PhD research on the internationalisation of telcos in the University of Adelaide. He also teaches international business and strategy MBA courses in the University of South Australia. In addition to his academic commitments, Mr Laanti offers consulting services in international business and strategy. Riku has a background in the ICT industry where he has worked for more than twelve years, including director positions in Telia Mobile/Telia Finland. Prior to his five-year career in Telia, he was a partner and managing director in one of the leading telecommunications retailers in Helsinki, Finland.

Pedro Lázaro received his MSc degree in telecommunication engineering in 1997 from the University of Málaga (Spain). From 1997 to 1999 he worked at the European Space Agency in Darmstadt (Germany). From 1999 to 2001 he worked at the International Telecommunication Union (ITU) in Geneva (Switzerland). Since 2001 he has been with the Communication Engineering Department, University of Málaga. His research interests include satellite and mobile communications, mainly focused on self-regulation of radio networks.

Heesang Lee is a professor of systems management engineering at Sungunkwan University in Seoul, Korea. He received a BS and anMS in industrial engineering from Seoul National University in 1983 and 1985, respectively. He got his PhD in Industrial & Systems Engineering from Georgia Institute of Technology, Atlanta, USA in 1991. Before joined as a faculty member of Sunkyunkwan University, he had been with KT and Hankuk University of Foreign Studies as a senior researcher and an associate professor, respectively. His current research interests include IT & telecommunication management, technology innovation, nature inspired modeling, and management of complex systems. His technical publications appear in international and domestic journals including *Management Science*, *European Journal of Operations Research*, and *Telecommunication Systems.*

Pierre Lévis received his degree from the Institut National des Telecommunications (INT) Evry France. He has been with France Telecom, R&D Division since 1998. He has worked in the specification, development, and evaluation of IP service offerings. His research interests are QoS, IPv6, network security, and mobile networks. He was an assistant professor at INT from 1990 to 1998, where he was in charge of computer networks courses, and his research interest was on information technology.

Chen Li is a post graduate student majored in information management and information system, School of Management and Economics, Beijing University of Posts and Telecommunications. His research interests focus on information integration, SOA and personalized recommendation system. Recently, the integration between BPM and SOA is being researched.

Yuanquan Li is a post graduate student majored in management science and engineering, School of Management and Economics, Beijing University of Posts and Telecommunications. His research interests focus on data mining for customer churn prediction, information system adoption and personalized recommendation system. Recently, service modeling for personalized recommendation in telecommunication industry is being researched.

Mirko Luca Lobina (m.lobina@diee.unica.it) is a researcher at the University of Cagliari (Italy) since 2001. His main research topics of interest are in multimedia networking (error recovery and concealment, IP telephony, and QoS in real-time services) and digital property protection mechanisms. He has published several journal articles and refereed conference papers. Dr. Lobina is also a freelance consultant for national and international IT enterprises.

Brian Low is a senior lecturer in marketing at the University of Western Sydney. Dr. Low has spent some sixteen years in the automotive, information technology and telecommunication sector, in various marketing and research positions. He has also consulted for a range of American, European and Chinese companies on marketing and internationalizing strategies and regulatory policies in the telecommunication sector in Asia Pacific. His research has been published in the *Journal of Business and Industrial Marketing, Industrial Marketing Management and the Journal of Asia – Pacific Business.* He is an occasional reviewer for the Journal of Telecommunications Policy.

Paolo Lupi works in the economic and competition analysis department of AGCOM, the Italian communication regulator, since 2001. He works both on research-oriented assignments and on-the-field projects. Main areas of expertise include competition analysis and regulatory policy (with a particular focus on fixed networks interconnection and Internet development). Paolo leads the team that is in charge of the analyses, both at retail and wholesale level, of the narrowband and broadband access markets. Paolo is a delegate to the Communication Infrastructures and Services Policy Working Group of the OECD and has participated (also as a chair) in the activities of many Working Groups of the European Regulators Group. Before joining AGCOM Paolo held research positions at the University of Naples (IT) and at the University of York (UK). He holds an MSc and a PhD in Economics from the University of York (UK). Paolo has published in academic and non academic journals on different topics related to the general areas of telecommunications economics and policy.

Maria Major, PhD (Manchester), is an assistant professor of management accounting at ISCTE – Business School, Lisbon. Her research interests lie in management accounting change both in private and public organisations. Dr. Major has published in academic journals including *Management Accounting Research* and *The European Accounting Review*. She is on the editorial board of *Journal of Accounting & Organizational Change* and she is additionally active in refereeing articles for a number of leading accounting journals. Currently she is conducting a research project funded by the Portuguese Government on management control in the healthcare sector.

Dimitrios G. Makris is an undergraduate student at the University of Peloponnese (UoP). His academic and research interests include mobile communications, network management as well as configuration and software engineering. He is a member of the "Communication Networks and Mobile Systems" laboratory of the University of Peloponnese (cnms.uop.gr) and participates in the "One Laptop Per Child" (OLPC) project.

Fabio M. Manenti, associate professor in Economics at the University of Padua (Italy). Ma in Economics at the University of Brescia (Italy), MSc and PhD in Economics at University of York (UK), Dottorato in Economia Politica at the University of Pavia (Italy). Research Interests include industrial organization, networks & Internet economics, software markets, economics of telecommunications, and regulation.

Sunilkumar S. Manvi completed his PhD from Indian Institute of Science, Bangalore, India. His areas of research include wireless/wired networks, AI applications in network management and E-commerce. He has published about 100 papers in journals/conferences. He has coauthored books *Communication Protocol Engineering* and *Computer Concepts and C Programming.* He is a reviewer of several international/national journals. He is a member of IEEE, USA, Fellow of IETE, India, Fellow of IE, India and member of ISTE, India. He has been included in Marqui's *Who's Who in World* and *International Biographies of Cambridge*, London in the year 2006.

Carol C. McDonough is a professor of economics at University of Massachusetts Lowell, in Lowell, Massachusetts. A member of the Zoning Board of Appeals in Andover, Mass., she holds an MA and a PhD in economics from Boston College. Dr. McDonough's primary area of teaching and research is microeconomics, especially applied microeconomics. She has written numerous articles and papers on the wireless telephony industry.

Fred McDougall holds a BComm(honours), MComm, and a Dip Ed from the University of Melbourne, a PhD from the University of Adelaide, and is a CPA (Fellow). He is currently the deputy vice-chancellor and vice-president (academic) at the University of Adelaide. He was previously the foundation chair of management, director of the graduate school of management, and executive dean of the faculty of the professions at the University of Adelaide.

James M. McKinion is a research electronics engineer with the USDA, Agricultural Research Service, Genetics and Precision Agriculture Research Unit located at Mississippi State, MS. Dr. McKinion has conducted research in cotton crop modeling, controlled-environment plant growth chambers, precision agriculture, and development of wireless networks for farm applications. In a 39 year research career with ARS, he has published over 200 papers in refereed journals, several book chapters and one book. He is a member of the American Society of Agricultural and Biological Engineers and an adjunct professor with the Agricultural and Biological Engineering Department, Mississippi State University.

John B. Meisel is a professor in the Department of Economics and Finance at Southern Illinois University Edwardsville. He received his PhD in economics at Boston College. He has published research examining the role of the Internet in transforming various communication and entertainment industries. Other research areas include antitrust, law and economics, and telecommunications.

Asunción Mochón received her MSc in economics from Autonoma University of Madrid and her PhD in economics from Universidad Nacional de Educación a Distancia (UNED). Currently, she is an associate professor of economics at Universidad Nacional de Educación a Distancia (UNED) in Madrid. Her research focuses on auctions in several markets, like spectrum licences. Most of her recent work has addressed key questions in multiunit auctions and combinatorial auctions. To this end, evolutionary computation techniques, like genetic algorithms, have been tested in several auctions models. She has published several articles on auctions in major journals and congress.

Pierrick Morand has been the project coordinator of the EU IST-MESCAL project. He also leads a research group within France Telecom working to develop and enhance VoIP services for corporate business. He graduated from the Ecole Nationale Supérieure d'Ingénieur de Caen (Institut des Sciences de la Matière et des Rayonnements), a French school of engineers, and received his degree from the University of Caen, France.

Frédéric Morlot was born on July 19th 1982 in Lyon (France). He majored in applied mathematics and pure mathematics in École Polytechnique and Telecom Paristech, both French universities. After his graduation, he worked in Orange Labs, the research and development division of France Telecom. At the same time, he is a PhD student in stochastic geometry at the École Normale Supérieure in Paris. His research fields include radio resource management and probabilities.

Álvaro Nascimento is the dean of the Faculty of Economics and Management at the Portuguese Catholic University, in Porto, Portugal, and has research and publications in the subjects of regulation of financial markets and telecommunications. He holds a PhD in Finance from Cass Business School (UK) and a MSc in International Trade and Finance from Lancaster University (UK). He was a visiting research fellow at the London Business School in the Regulation Initiative and a consultant to several financial and telecommunication companies on regulatory and competition issues. He was also the lead economist on competition issues for recent M&A operations in Portugal (representing the companies).

Ricardo Neisse is a final year PhD candidate at the University of Twente, the Netherlands. He received his MSc degree in 2004 from the Federal University of Rio Grande do Sul, Brazil. His broad research area is management of distributed systems and networks, and the topic of his PhD thesis is security and trust management for context-aware service platforms. He has published several papers and actively contributed as a member of the scientific staff in many international conferences and research projects.

Phillip Olla is the endowed Phillips chair of management and professor of MIS at the school of business at Madonna University in Michigan USA, and he is also a visiting research fellow at Brunel University, London, UK. His research interests include knowledge management, mobile telecommunication, and health informatics. In addition to university level teaching, Dr. Phillip Olla is also a chartered engineer and has over 10 years experience as an independent consultant and has worked in the telecommunications, space, financial and healthcare sectors. He was contracted to perform a variety of roles including chief technical architect, program manager, and director. Dr Olla is the associate editor for the *Journal of Information Technology Research and the Software* / book review editor for the

International Journal of Healthcare Information Systems and Informatics, and is also a member of the editorial advisory & review board for the *Journal of Knowledge Management Practice*. Dr. Phillip Olla has a PhD in mobile telecommunications from Brunel University in the UK. He is an accredited press member of the British Association of Journalism, chartered IT professional with the British Computing Society and a member of the IEEE society.

Claudia Parlanti is currently working at Procter and Gamble Italy (Rome) in the Marketing Department. She received her master's of science in economics and management of international markets and new technologies from Bocconi University, Milan, in 2007 and her degree in economics of international markets and new technologies from Bocconi University, Milan, in 2005. In 2007, she worked for six months at Cisco System Italy (Milan) as a junior analyst. Her main interests include product innovations, technology diffusion and adoption in high tech sectors.

Jong-Hyun Park is a senior researcher in technology strategy research Division of Electronics and Telecommunications Research Institute (ETRI). He received the BA degree in economics from Korea University, Seoul, Korea and the MS degree in management from information and communications University (ICU), Daejeon, Korea. He is mainly interested in IT market analysis, marketing strategies and ubiquitous network.

Myeong-Cheol Park received the BS degree in the industrial engineering and MS in business administration from Seoul National University, Seoul, Korea. He received his PhD degree in management science from University of Iowa. He worked for ETRI, where he engaged in many research related in IT management and economic issues. He is professor at the school of IT business, Information and Communications University (ICU). He main research interests include IT management and strategies.

Harry G. Perros is a professor of computer science, an alumni distinguished graduate professor, and the program coordinator of the Master of Science degree in computer networks at NC State University. He has published extensively in the area of performance modeling of computer and communication systems, and he has organized several national and international conferences. He has also published three print books and an e-book. His current research interests are in the areas of next generation networks, multi-domain routing, resource allocation under QoS, service science management and engineering (SSME), and queuing theory.

Samuel Pierre is currently a professor of computer engineering at École Polytechnique de Montréal where he is director of the Mobile Computing and Networking Research Laboratory (LARIM) and NSERC/Ericsson Industrial Research Chair in Next-generation Mobile Networking Systems. Dr. Pierre is the author or co-author of 6 books, 19 book chapters, 17 edited books, as well as more than 350 other technical publications including journal and proceedings papers.

James Prieger is an economist at the Pepperdine University School of Public Policy, where he is an associate professor. He specializes in regulatory economics, industrial organization, and applied econometrics. His research includes the impact of regulation on telecommunications innovation, the use of cell phones while driving, the broadband digital divide, the impact of the Americans With Disabilities Act on retail firms, and econometric methodology for non-randomly sampled duration data. Prieger sits on the editorial board of *Applied Economics Quarterly*. He received his BA from Yale University and his PhD from the University of California, Berkeley.

Jiayin Qi is an associate professor at the School of Economics & Management, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include customer relationship management and telecommunication operation management. She has published more than 60 papers and 3 books in the two research fields.

Sally Rao (hons) PhD CPM is a senior lecturer in marketing at the Business School, University of Adelaide. She is the lecturer in charge of consumer behaviour and e-marketing. Her research interests include relationship marketing, internet marketing, services marketing and technology adoption. She is an active researcher and has published in the *Journal of Business and Industrial Marketing, European Journal of Marketing, Qualitative Market research: an International Journal, Australasian Marketing Journal, International Journal of Internet Marketing and Advertising and Journal of Internet Business.* She has won the best paper award in an international conference.

Salah Ramadan (smageid@engr.smu.edu) received hisMS and PhD in systems and computers engineering from Al-Azhar University in 2002 and 2005, respectively. Since 2005, he is a lecturer in systems and computers engineering at Al-Azhar University and an instructor in CISCO academy, Egypt. Currently, he is a research scholar in the Computer Science and Engineering Department, School of Engineering, the Southern Methodist University at Dallas, TX, USA, where he works in post doctoral research. His research interests include mobile computing, cellular networks, and sensor networks.

James Salter completed his PhD in the field of optimisation of P2P networks at the Department of Computing, University of Surrey, UK in September 2006. He also holds a BSc in computing and information technology (1st class) from the same University in 2003. He has published numerous research articles in fully refereed journals and international conferences in the fields of peer-to-peer networks and agent-based recommender systems and he has received a number of best paper awards.

Yeong-Wha Sawng is a senior researcher of the IT Technology Strategy Research Division at the Electronics and Telecommunications Research Institute (ETRI), Korea. He received an MIM in Management from Whitworth University, USA, and a PhD in information technology management from Hanyang University, Seoul, Korea in 1995 and 2006, respectively. He joined ETRI in 2000, and has been working in the areas of digital convergence, IT policy, mobile application, technology management and business strategy. His research interests also include high-tech marketing, technology management strategy, business model, and consumer behavior. His academic papers were published in several international and Korean journals.

Carsten Schröer is a doctoral candidate in the E-Business and E-Entrepreneurship Research Group at the University of Duisburg-Essen, Germany. He received a degree in business administration and international management from the Dresden University of Technology, Germany and holds an MBA from Pfeiffer University at Charlotte, NC, USA. His research is currently focused on business angels, the informal capital market, and e-business.

George Scheets received a BS in general engineering from the United States Military Academy at West Point in 1975, and following five years in the U.S. Army as a Signal Corps officer,MS and PhD degrees in electrical engineering from Kansas State University in 1984 and 1987 respectively. He is an associate professor at Oklahoma State University. His research interests include telecommunications network analysis and design, teaching in a virtual environment, and secure communications protocols. He has performed contract research for organizations such as Williams Telecommunications, the U.S. Air Force, Motorola Government Electronics Group, the U.S. Department of Education, and the U.S. Department of Defense.

Dong-Hee Shin has been an assistant professor in the School of Information Sciences and Technology at Penn State University since 2004. He earned his PhD andMS in information sciences and telecommunications from Syracuse University. He has been researching on telecommunications management and policy and he had various consultancy experiences in the field of telecommunications and information industry. His current project looks at the consequences of governmental policy and regulations on telecommunications and information infrastructure as well as the applications and implications of these services. He is also interested in evaluating the process and impact of large technological projects and policy by using qualitative and quantitative case study methods. His publications have appeared in several international journals such as *Information Research*, *Technological Forecasting & Social Change*, *INFO*, and *Telematics and Informatics.*

Yajing Si is a lecture professor at the school of Logistics, Beijing Wuzi University, Beijing, China. Her research interests include the customer relationship management and telecommunication operation management. She is also a PhD candidate at Beijing University of Posts and Telecommunications.

Darko Skorin-Kapov is a professor of information technology and operations management at the School of Business, Adelphi University, New York. He received his PhD in management science from the University of British Columbia, Vancouver, Canada, and MSc in information science and BSc in mathematics from Zagreb University, Croatia. He authored numerous research papers in management science and related areas. His research interests are focusing on optimization and cost allocation in communication networks. Dr. Skorin-Kapov is a recipient of a number of awards and grants, including Bender Award, Fulbright Award, and National Science Foundation Grants.

Jadranka Skorin-Kapov is a professor of operations management at the College of Business, Stony Brook University. She received a PhD in operations research from the University of British Columbia in Canada, a PhD and MA degrees in Philosophy from Stony Brook University in USA, and a MSc and BSc degrees in mathematics from the University of Zagreb in Croatia. Her research includes applications of discrete optimization to telecommunications, network design, manufacturing design, scheduling, and location and layout. She authored many research papers and is a recipient of a number of grants, including six National Science Foundation grants.

Marc St-Hilaire received the computer engineering degree from Concordia University in 2002 and the master and PhD degrees in computer engineering respectively in 2003 and 2006 from the École Polytechnique de Montréal. Dr. St-Hilaire is currently an assistant professor in the Department of Systems and computer engineering at Carleton University. His research interests include telecommunication network planning, mathematical modeling and meta-heuristics.

Christoph Stöckmann is a doctoral candidate in the E-Business and E-Entrepreneurship Research Group at the University of Duisburg-Essen, Germany. He received a degree in business informatics from the University of Duisburg-Essen, Germany. His research is currently focused on strategic management, corporate entrepreneurship, entrepreneurial orientation, ambidextrous innovation, and e-business.

Timothy S. Sullivan is an instructor in the Department of Economic and Finance at Southern Illinois University Edwardsville. He received his PhD in economics at the University of Maryland. He has published research examining the role of Internet portals and the Internet's effect on the music and motion picture industries. Other research areas include local economic development and statistical methods.

Csaba A. Szabó received his PhD degree from the Budapest University of Technology and Economics (BME) and a Doctor of Technical Sciences title from the Hungarian Academy of Sciences. He is a Professor at the Dept. of Telecommunications of BME. He has been also with Create-Net, an international research center, based in Trento, Italy, as a Senior Advisor since 2004. His 30+ years of experience in academia, R&D and telecommunication business includes metropolitan area networks, integrated services wireless networks, video conferencing and media streaming, broadband optical and wireless networks, community network technologies and applications. Dr. Szabó has been a member of editorial boards of several journals including *Computer Networks*, and is the editor-in-chief of the Hungarian journal *Telecommunications*. He has been general chair and steering committee chair of several international conferences, including Multimedia Services Access Networks (MSAN), the 1st Int'l IEEE/Create-Net Workshop on Telemedicine over Broadband (BroadMed), and the series of IEEE/Create-Net Conference on Testbeds and Research Infrastructures (Tridentcom). Dr. Szabó is a co-editor and co-author of a recent Wiley book titled *Broadband Services: Business Models and Technologies for Community Networks*. He is a senior member of IEEE.

Jing Tan is a post graduate student in the School of Economics and Management of Beijing University of Posts and Telecommunications (BUPT), majoring in the Management Science and Engineering. She graduated from BUPT and received her bachelor's degree of business administration in July, 2007. Her research interests focus on the Customer Relationship Management (CRM), especially in the telecommunication industry.

Katalin Tarnay is professor emeritus at the University of Pannonia (Veszprém, Hungary). She received MSc degree from the Technical University of Budapest in 1956, PhD degree from the Hungarian Academy of Sciences in 1974. She developed software tools and methods for protocol engineering first in the Central Research Institute for Physics, and later in NOKIA Hungary. At the same time she was appointed full professor of the University of Pannonia and honorary professor of the Technical University of Budapest. She was author of the book *Protocol Specification and Testing* (Plenum Press, New York, 1991.)

Indrit Troshani PhD, MSc (computer-based InfoSys), GradCertEd (TertTeach), BBA (hons), MACS, teaches Internet commerce, electronic commerce, and information systems at both undergraduate and postgraduate levels at the Business School, University of Adelaide. His research interests include adoption and diffusion of network innovations (e.g. XBRL) and mobile services (e.g. 3G). Indrit's work has appeared in *European Journal of Innovation Management*, *Electronic Markets*, *International Journal* *of E-Business Research, Journal of Theoretical and Applied Electronic Commerce Research* and others. Indrit is a member of Australian Computer Society.

Nikolaos D. Tselikas is a research associate in the School of Electrical and Computer Engineering of National Technical University of Athens (NTUA) and a visiting lecturer in University of Peloponnese (UoP). He is a reviewer in several IEEE and Elsevier journals and his research interests include service-oriented architectures, service creation and control, context-aware services, location based services, distributed systems as well as middleware and software engineering. He received both his Dipl.-Ing. degree and his PhD from the School of Electrical and Computer Engineering of NTUA in 1999 and 2004, respectively.

Belén Usero is an assistant professor of Strategic Management at University Carlos III de Madrid (Spain). She holds a PhD in economy and a Licentiate degree in Business Administration (University Carlos III de Madrid). Previously, she has taught strategic management at University of Salamanca. Her main research interests include first- mover advantages in mobile telecommunications industry, corporate strategy of SME, non-market strategy and business models. She has published her research in national and international journals such as *Service Business* — an international journal

Dimitris Varoutas holds a physics degree and MSc and PhD diplomas in communications and technoeconomics from the University of Athens. Since 2005, he is a lecturer on telecommunications technoeconomics in the Department of Informatics and Telecommunications at the same university. He actively participates in national and EU-funded R&D technoeconomic activities for telecommunications, networks and services such the ICT-OMEGA and the CELTIC/CINEMA projects, as well as the Conferences on Telecommunications TechnoEconomics. His research interests include optical and wireless communications systems, technoeconomic evaluation of network architectures and services and demand and price modeling. He has published more than 40 publications in refereed journals and conferences in these topics and he is a member of several IEEE societies, serving as a reviewer.

Iakovos S. Venieris is a professor in the School of Electrical and Computer Engineering of NTUA. His research interests are in the fields of broadband communications, Internet, mobile networks, intelligent networks, internetworking, signaling, service creation and control, distributed processing, agents technology, and performance evaluation. He received his Dipl.-Ing. degree from the University of Patras in 1988, and his PhD degree from NTUA in 1990, all in electrical and computer engineering. He is a reviewer for several IEEE, ACM, Elsevier, and John Wiley journals, associate editor for the IEEE Communications Letters, member of the editorial staff of Computer Communications (Elsevier), and has been guest editor in the *IEEE Communications Magazine*. Prof. Venieris is the co-editor of two international books on *Intelligent Broadband Networks* (Wiley 1998) and *Object oriented Software Technologies in Telecommunications* (Wiley 2000).

Hannu Verkasalo acts as a project manager of the national research project MoMI (Modeling the Mobile Internet Usage and Business) at the Helsinki University of Technology in Finland. Hannu Verkasalo has actively published on the evolution of the mobile services market. He is currently finishing his doctoral dissertation on the emergence of the mobile Internet and associated empirical research methods. Verkasalo has particularly focused on the empirical modeling of mobile services, utilizing the recently developed handset-based research method. In addition to research work and consulting in the mobile industry, Verkasalo gives lectures on network economics, operator business and telecommunication regulation.

Ricardo Lemos Vianna received his BSc and MSc degrees, both in computer science, from the Federal University of Rio Grande do Sul (UFRGS) in 2003 and 2007, respectively. His main topic of interest is the use of Web services technologies, especially service composition using WSBPEL, applied to computer network management.

Barney Warf is professor of geography at the University of Kansas. His research and teaching interests lie within the broad domain of human geography, particularly economic and political issues. His work straddles contemporary political economy and social theory on the one hand and traditional quantitative, empirical approaches on the other. His work focuses primarily on the dynamics of the service economy, including the telecommunications industry, New York as a global city, electronic financial markets, offshore banking, international legal and engineering services, the geographies of cyberspace, the regional impacts of military spending, political geography, and time-space compression. He is also interested in different perspectives on the philosophy of space and time. He has authored, co-authored or co-edited seven books, 25 book chapters, and 85 journal articles, and won teaching and research awards.

Ran Wei (PhD, Indiana University, 1995) is an associate professor in School of Journalism & Mass Communications at the University of South Carolina, USA. His research interest focuses on the adoption, use and impact of new media technology in national, organizational, and individual contexts. He is a renowned scholar in the emerging field of mobile communication. His publications include 40 refereed journal articles and book chapters, and one co-authored book.

James G. Williams received his PhD from the University of Pittsburgh in Information and computer science in 1972. After teaching at the University of South Caroline from 1972 to 1974 he returned to the University of Pittsburgh's Department of Information Science as an assistant professor in 1974 and was promoted to full professor in 1984. He served as department chairman from 1985 to 1993. He retired from the University of Pittsburgh as a Professor Emeritus to become CIO of Broadstreet Communications, Inc. and later become president of Automated Systems Research and Development, Inc., a position he currently holds. During his academic career he published nearly 100 journal and proceedings articles, co-authored 2 encyclopedias with Allen Kent, and co-authored 5 books and received millions of dollars in research grants.

Kaiqi Xiong is currently on the faculty of the Department of Computer Science at Texas A&M University, Commerce, Texas, United States. He has received the PhD degree in computer science and theMS degree in computer engineering from the North Carolina State University, and the PhD degree in mathematics from the Claremont Graduate University. He has served as a TPC member for a variety of international conferences, such as IEEE ICC'07, IEEE Globecom'08, CSA'08, and IEEE ICC'09. His current research interests are in the areas of computer and network security, distributed service computing, performance evaluation, and next generation networks

Zaharias D. Zaharis received the BSc degree in physics in 1987, the MSc degree in electronics in 1994 and the PhD degree in physics in 2000, all from the Aristotle University of Thessaloniki, Greece. Since 1993 he has been working as a research assistant in the Radiocommunications Laboratory at the Department of Physics, Aristotle University of Thessaloniki, Greece. Since 2002 he also has been working in the Telecommunications Center of the Aristotle University of Thessaloniki, Greece. His research interests include design and optimization of antenna arrays, design and optimization of microwave circuits and systems, mobile communications systems, RF measurements, wave propagation, electromagnetic scattering and computational electromagnetics.

Mohamed Zaki is the professor of software engineering, Computer and System Engineering Department, Faculty of Engineering, Al-Azhar University at Cairo. He received his BSc and MSc degrees in electrical engineering from Cairo University in 1968 and 1973 respectively. He received his Ph. D. degrees in computer engineering from Warsaw Technical University, Poland in 1977. His fields of interest include artificial intelligence, soft computing, and distributed systems.

Yangming Zhang received his bachelor's degree of management information systems and Master degree of management engineering from Beijing University of Posts and Telecommunications (BUPT), China, in 2000 and 2003, respectively. He is currently pursuing his PhD degree of management science in BUPT. His research interests include customer relationship management (CRM) and business intelligence.

Yingying Zhang received both her bachelor's degree and master's degree of information management and information systems from Beijing University of Posts and Telecommunications (BUPT), China, in 2004 and 2008 respectively. During her undergraduate period, she also minored in telecommunication engineering. Her research interests include customer satisfaction measurement, customer relationship management and customer behavior analysis.

Zizhong Zhao (PhD, Communication University of China, 2004) is associate professor and director of The Center for New Media Studies at the Communication University of China, Beijing, China. His research interests include strategies of radio, film and TV industries in China, media convergence, and developments of a Cultural Media Recreation Park in Beijing. He is a consultant to China's telecommunications policymakers and several Chinese cable system operators. He has published four books.

Index

C

D

E

F

G

H

N

O

P

Q

R

S

T

U

V

W

Y

Z