



# Advances in Intelligent and Soft Computing

## Editor-in-Chief

Prof. Janusz Kacprzyk  
Systems Research Institute  
Polish Academy of Sciences  
ul. Newelska 6  
01-447 Warsaw  
Poland  
E-mail: kacprzyk@ibspan.waw.pl

---

Further volumes of this series can be found on our homepage: [springer.com](http://springer.com)

Vol. 112. L. Jiang (Ed.)  
*Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19-20, 2011, Melbourne, Australia, 2011*  
ISBN 978-3-642-25193-1

Vol. 113. J. Altmann, U. Baumöl, and B.J. Krämer (Eds.)  
*Advances in Collective Intelligence 2011, 2011*  
ISBN 978-3-642-25320-1

Vol. 114. Y. Wu (Ed.)  
*Software Engineering and Knowledge Engineering: Theory and Practice, 2011*  
ISBN 978-3-642-03717-7

Vol. 115. Y. Wu (Ed.)  
*Software Engineering and Knowledge Engineering: Theory and Practice, 2011*  
ISBN 978-3-642-03717-7

Vol. 116. Yanwen Wu (Ed.)  
*Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), 2012*  
ISBN 978-3-642-11275-1

Vol. 117. Yanwen Wu (Ed.)  
*Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), 2012*  
ISBN 978-3-642-25436-9

Vol. 118. A. Kapeczynski, E. Tkacz, and M. Rostanski (Eds.)  
*Internet - Technical Developments and Applications 2, 2011*  
ISBN 978-3-642-25354-6

Vol. 119. Tianbiao Zhang (Ed.)  
*Future Computer, Communication, Control and Automation, 2011*  
ISBN 978-3-642-25537-3

Vol. 120. Nicolas Loménie, Daniel Racoceanu, and Alexandre Gouaillard (Eds.)  
*Advances in Bio-Imaging: From Physics to Signal Understanding Issues, 2011*  
ISBN 978-3-642-25546-5

Vol. 121. Tomasz Traczyk and Mariusz Kaleta (Eds.)  
*Modeling Multi-commodity Trade: Information Exchange Methods, 2011*  
ISBN 978-3-642-25648-6

Vol. 122. Yinglin Wang and Tianrui Li (Eds.)  
*Foundations of Intelligent Systems, 2011*  
ISBN 978-3-642-25663-9

Vol. 123. Yinglin Wang and Tianrui Li (Eds.)  
*Knowledge Engineering and Management, 2011*  
ISBN 978-3-642-25660-8

Vol. 124. Yinglin Wang and Tianrui Li (Eds.)  
*Practical Applications of Intelligent Systems, 2011*  
ISBN 978-3-642-25657-8

Vol. 125. Tianbiao Zhang (Ed.)  
*Mechanical Engineering and Technology, 2011*  
ISBN 978-3-642-27328-5

Vol. 126. Khine Soe Thuang (Ed.)  
*Advanced Information Technology in Education, 2011*  
ISBN 978-3-642-25907-4

Vol. 127. Tianbiao Zhang (Ed.)  
*Instrumentation, Measurement, Circuits and Systems, 2012*  
ISBN 978-3-642-27333-9



Tianbiao Zhang (Ed.)

---

# Instrumentation, Measurement, Circuits and Systems

## Editor

Tianbiao Zhang  
Huazhong Normal University,  
China  
E-mail: ztb@mails.ccnu.edu.cn

ISBN 978-3-642-27333-9

e-ISBN 978-3-642-27334-6

DOI 10.1007/978-3-642-27334-6

Advances in Intelligent and Soft Computing

ISSN 1867-5662

Library of Congress Control Number: 2011943814

© 2012 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Typeset* by Scientific Publishing Services Pvt. Ltd., Chennai, India

Printed on acid-free paper

5 4 3 2 1 0

springer.com

# ICIMCS 2011 Preface

2011 International Conference on Instrumentation, Measurement, Circuits and Systems (ICIMCS 2011) will be held on December 12–13, 2011, Hong Kong.

Instrumentation is defined as the art and science of measurement and control of systems. An instrument is a device that measures or regulates physical quantity/process variables such as flow, temperature, level, or pressure. Instruments include many varied contrivances that can be as simple as valves and transmitters, and as complex as analyzers. Instruments often comprise control systems of varied processes such as refineries, factories, and vehicles. The control of processes is one of the main branches of applied instrumentation. Instrumentation can also refer to handheld devices that measure some desired variable. Diverse handheld instrumentation is common in laboratories, but can be found in the household as well.

Measurement is the process or the result of determining the ratio of a physical quantity, such as a length or a mass, to a unit of measurement, such as the meter or the kilogram. The science of measurement is called metrology.

The research area of the Circuits and Systems covers the theory and applications of circuits and systems, signal processing, and VLSI circuit and system design methodology.

ICIMCS 2011 will be the most comprehensive Conference focused on the various aspects of advances in Instrumentation, Measurement, Circuits and Systems. Our Conference provides a chance for academic and industry professionals to discuss recent progress in the area of Instrumentation, Measurement, Circuits and Systems. The goal of this Conference is to bring together the researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to the multifaceted aspects of Instrumentation, Measurement, Circuits and Systems.

To our satisfaction, our conference encompasses a large number of research topics and applications: from Instrumentation to Network Theory and Applications; from Human-Machine Systems and Cybernetics to Geometric modeling and Fractals and other related topics are included in the scope of this conference. In order to ensure high-quality of our international conference, we have high-quality reviewing course, each paper will be reviewed by at least two reviewers, who are experts and professors from home and abroad and low-quality papers have been refused. All accepted papers will be published by Springer.

We are proud to invite Professor Jun Wong, a Professor and the Director of the Computational Intelligence Laboratory in the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong, as our Keynote speaker. Therefore, I am sure you will gain more information than expected from the proceedings.

I very sincerely thank the authors who have contributed to the proceedings and the referees who reviewed the quality of the submitted contributions. Our sponsors' support, either financial or moral, is gratefully acknowledged. The members of the Organizing Committee as well as other numerous individuals behind the scenes are greatly appreciated for their tireless effort and dedication in the organization of the conference.

Tianbiao Zhang  
Publication Chair

# ICIMCS 2011 Organizer

## Honor Chair

David Wang                      IEEE Nanotechnology Council Cambodia Chapter  
Past Chair, Cambodia

## General Chairs

Mark Zhou                      Hong Kong Education Society, Hong Kong

## Publication Chair

Tianbiao Zhang                Huazhong Normal University, China

## Organizing Chairs

Khine Soe Thaug                Maldives College of Higher Education, Maldives  
Biswanath Vokkarane        Society on Social Implications of Technology  
and Engineering

## International Committee

Yiyi Zhouzhou                Azerbaijan State Oil Academy, Azerbaijan  
Ying Zhang                      Wuhan University, China  
David Wang                      IEEE Nanotechnology Council Cambodia Chapter Chair,  
Cambodia  
Srinivas Aluru                ACM NUS Singapore Chapter, Singapore  
Tatsuya Akutsu                ACM NUS Singapore Chapter, Singapore  
Aijun An                        National University of Singapore, Singapore  
Qinyuan Zhou                Jiangsu Teachers University of Technology, China  
Mark Zhou                      Hong Kong Education Society, Hong Kong  
Yiyi Zhouzhou                Azerbaijan State Oil Academy, Azerbaijan  
Tianbiao Zhang                Huazhong Normal University, China  
Garry Zhu                        Thompson Rivers University, Canada  
Ying Zhang                      Wuhan University, China

# Contents

<b>Range of Query Processing in Peer to Peer Networks</b> . . . . .	1
<i>R. Saravanan, P. Vivekananth</i>	
<b>A Scene Detection Based Scheme for Supporting VCR Functions</b> . . . . .	11
<i>Shu-bin Xu</i>	
<b>Constructing Digital Campus Using Campus Smart Card System</b> . . . . .	19
<i>Fang Wang, Zhuo-sheng Jia</i>	
<b>Research about Pattern Matching Algorithm</b> . . . . .	27
<i>Minjie Wang, Lianxuan Zhu</i>	
<b>Study of E-mail Filtering Based on Mutual Information Text Feature Selection Method</b> . . . . .	33
<i>Shangfu Gong, Xingyu Gong, Yuan Wang</i>	
<b>Research on the Moving Vehicle Detection Algorithm Based on the Motion Vector</b> . . . . .	41
<i>Jian Gong, Fuqiang Liu, Chunlin Song, Jianzhu Cui, Zhipeng Li</i>	
<b>A Robust Blind Watermarking of Vector Map</b> . . . . .	51
<i>Lian-quan Min, Xian-zhen Zhu, Qiang Li</i>	
<b>Implementation of Spectrum Sensing Based on Covariance in Cognitive Radio</b> . . . . .	61
<i>Hongyu Li, Yuchen Wang, Yading Chen, Shaoqian Li</i>	
<b>Simulation and Analysis of Quiet DDOS Attacks</b> . . . . .	71
<i>Jing Zhang, Bo Liu, Huaping Hu, Lin Chen, Tianzuo Wang</i>	
<b>VV&amp;A Resources Management System of Guidance Simulation</b> . . . . .	83
<i>Zhenjun Li, Peng Jiao</i>	

<b>Web-Based Collaborative Construction of Curriculum Texts in Higher Education</b> .....	89
<i>Wenxin Deng, Fang Lu</i>	
<b>Study on LDPC Application in Wireless Communication System</b> .....	99
<i>Li Wang, Tikui Wang, Jianqiang Guo, Jinlong Li, Xiaorong Gao</i>	
<b>Coding of Magnetic Ruler in Cylinder</b> .....	105
<i>Gengbiao Chen, Yang Yang</i>	
<b>Edge Detection of Level Ruler with Digital Level</b> .....	111
<i>Gengbiao Chen, Caiyi Wang</i>	
<b>New Approaches to Apply Multi-stereo Matching into 3DCG Modeling</b> .....	117
<i>Ningping Sun, Hiroshi Nagaoka</i>	
<b>Research on Sleep Mode in WANDAS</b> .....	127
<i>Li Meng, Su Han, Kang Yu, Yun Chao</i>	
<b>Design and Realization of Integrated Ultrasound Transmitting Circuit</b> .....	137
<i>Hao Liu, Changyun Miao, Yongmao Feng, Feng Rong</i>	
<b>Special Way of Optical Communications the Technique of Superluminal Communications</b> .....	145
<i>Zi Hua Zhang</i>	
<b>Design and Process Simulation on High Voltage VDMOS</b> .....	155
<i>Shouguo Zheng, Jian Zhang, Xinhua Zeng, Zelin Hu, Zede Zhu</i>	
<b>Parameters Estimation of Stratospheric Airship and Model of Inflating and Deflating Air</b> .....	165
<i>Shaoxing Hu, Erbing Wu, Aiwu Zhang</i>	
<b>A New Method for Improving Life Time in Wireless Sensor Network by Using Genetic Algorithm</b> .....	177
<i>Shahram Babaie, Arash Khadem</i>	
<b>Modern Education Technology and the Transformation of Teacher Role</b> .....	185
<i>Canlong Wu, Shuying Zhan</i>	
<b>Network Security Risk Assessment Model and Method Based on Situation Awareness and CORAS</b> .....	191
<i>Yong Qi, Yan Wang, Qianmu Li</i>	
<b>Unit Transient Based Protection for High Voltage Wind Power Transmission Lines</b> .....	205
<i>Mingchao Xia, Yong Lin, Edward W.C. Lo</i>	
<b>Assessment of EXPO's Impact on Regional Economy</b> .....	219
<i>Shiyun Yu, Hao Li</i>	

<b>An Optimal Model for Repeater Coordination</b> .....	225
<i>Shiyun Yu, Hao Li</i>	
<b>Copyright Self-registration and Its Secure Authentication Based on Digital Watermark</b> .....	233
<i>Jianbo Liu, Cheng Yang</i>	
<b>A New Approach for SQL-Injection Detection</b> .....	245
<i>Cong-cong Shi, Tao Zhang, Yong Yu, Weimin Lin</i>	
<b>Cloud Computing Concept in the Application of HIT Library</b> .....	255
<i>Xiaodan Wang, Tingbo Fu, Guang Yu, Jiaomei Sun</i>	
<b>Customer Segmentation and Marketing Strategy of Commercial Banks Based on CLV</b> .....	261
<i>Xia Wang, Lan Liu</i>	
<b>Design and Implementation of Media Independent Handover on Heterogeneous Wireless Networks</b> .....	269
<i>Chang-Yi Kao, Chin-Shyurng Fahn, Jia-Liang Tsai</i>	
<b>On-Line Detection and Evaluation of Cavitation in Large Kaplan Turbines Based on Sound Wave</b> .....	277
<i>Huixuan Shi, Xuezheng Chu, Zhaohui Li</i>	
<b>Network Capacity Analysis for Survivable WDM Optical Networks</b> .....	291
<i>Zhenrong Zhang, Zhengbin Li, Yongqi He</i>	
<b>Study on Traffic Information Dissemination Scheme in Transit of Highway Network</b> .....	297
<i>Jun-jun Tang</i>	
<b>Analysis and Emulation of a Layered Aggregation Method Based on Fair</b> .....	305
<i>Youjun Chen, Yunbing Zeng, Hongying He</i>	
<b>MIP: A New Routing Metric for Wireless Mesh Networks</b> .....	313
<i>Sufen Zhao</i>	
<b>A Wireless Coordinating Scheme Proposed to Improve the Dispatching Security in China's Railway System</b> .....	321
<i>Junfeng Lu, Wen-qiang Lu</i>	
<b>The Prospects and Development on FTTH Technology in China</b> .....	329
<i>Jin Cai, Zhiyong Tao, Huijuan Zhang, Lu Qin, Qian Mao</i>	
<b>Cooperative Behavior in Evolutionary Snowdrift Game on a Growing Complex Network</b> .....	335
<i>Qingke Zeng, Haiyong Yang, Hong Wang</i>	



<b>Research and Implementation on a Collaborative Style of Test Bank System</b> .....	341
<i>Yi-Bing Liu</i>	
<b>Thermal Power Plant Pressure Dense Phase Pneumatic Ash PLC Control System Design and Implementation</b> .....	347
<i>Binlin Li, Wen You, Dianbin Song, Quan Tao</i>	
<b>Communication Design for I2C Bus Based on Proteus</b> .....	355
<i>Fengjie Wu, Liming Wu, Tangmei He</i>	
<b>Application of Proteus in Microcontroller Comprehensive Design Projects</b> .....	363
<i>Fengjie Wu, Tangmei He</i>	
<b>Modelling Three-Dimensional Foundation Pit Using Python Scripts</b> .....	371
<i>Shi-lun Feng, Yong Han, Pu-lin Li, Jun Li</i>	
<b>The Research of the Cloud Security Architecture</b> .....	379
<i>Hu Xiangyi, Ma Zhanguo, Liu Yu</i>	
<b>The Energy Harvesting Tipping Point for Wireless Sensor Applications</b> . . . .	387
<i>Alan Pang, Ross Bannatyne</i>	
<b>Design of Broadband Optical Access Network Based on 10G-EPON</b> .....	393
<i>Duo Peng, Jianbiao Wang</i>	
<b>Comprehending Collective Wisdom in Hall for Workshop of Metasynthetic Engineering Based on Semantic Web</b> .....	401
<i>Sao-qi Zhou, Ren Deng, You-hong Fan</i>	
<b>A General Result on the Convergence of Online BP Learning</b> .....	409
<i>Rui Zhang, Zuo-Zhi Liu</i>	
<b>Research and Application of Data Synchronization Based on Group Communication Technology</b> .....	417
<i>Li-xin Li, Jue Wang</i>	
<b>Kinetic Study of Vanadium-Oxides' Magneto-sputter Deposition Based on Experimental Reaction Data</b> .....	427
<i>Zheng Cui, Yu Lu, Zuochun Shen, Jianye Lu</i>	
<b>Cost Forecast of Overhead Transmission Line Based on BP Neural Network</b> .....	433
<i>Jun Zhou, Hanliang Wang, Yueguang Yang, Jiangjun Ruan</i>	
<b>Dynamic Aggregation Method of Induction Motors Based on Coherent Characteristics</b> .....	441
<i>Jun Zhou, Jian Zhang, Yueguang Yang, Yifeng Ju</i>	

<b>Application of the Monopulse Tracking in Mimo Radar</b> .....	457
<i>Guodong Qin, Baixiao Chen</i>	
<b>A Class of Differential 4 Uniform Functions from Gold Functions</b> .....	467
<i>Dabin Zheng</i>	
<b>The Stability of USO to Mars Orbiter Influence</b> .....	477
<i>Zhen Wang, Kun Shang, Nan Ye, Na Wang, Jin Song Ping</i>	
<b>Review of the Check Method in Metrology Support for Equipment</b> .....	485
<i>Wei Jiang</i>	
<b>Fuzzy Control of Denitrifying Phosphorus Removal via Nitrite</b> .....	493
<i>Bin Ma, Yongzhen Peng, Shuying Wang, Qing Yang, Xiyao Li, Shujun Zhang</i>	
<b>Application and Challenge of It Service Management to University Informatization</b> .....	501
<i>Hong Lin, Yajuan Sun</i>	
<b>A Novel Method for Iris Recognition Based on Zero-Crossing Detection</b> . . . .	509
<i>Li Yu, Xue Zhou</i>	
<b>Fast Iris Location Based on Window Mapping Method</b> .....	519
<i>Li Yu, Xue Zhou</i>	
<b>The Principles of Disk Taint Analysis of Temu and How to Write a Plugin</b> .....	527
<i>Lanlan Qi, Jiangtao Wen, Bing Zhou, Yu Chen, Zhiyong Wu</i>	
<b>Study for the Cooperation of Commercial Procedure Basing on Agent Technique</b> .....	535
<i>Shiyong Ding, Shuli Wang</i>	
<b>An Analysis and Assessment Method of WCDMA Network Traffic</b> .....	541
<i>Xinzhou Cheng, Jian Guan</i>	
<b>Research on Embedded Parallel System Based on Ethernet</b> .....	547
<i>Li Ying, Li Bo</i>	
<b>A PMAC Motor Drive with a Minimum Number of Sensors</b> .....	557
<i>Li Ying, Nesimi Ertugrul</i>	
<b>The Integrated Concept of the Ground Segment for Space Internet</b> .....	569
<i>Jian-ping Liu, Chun-yan Luo, Xiao-ying Li, Fan Yang</i>	
<b>Related-Mode Attacks on PMAC</b> .....	577
<i>Jin Xu, Qiaoyan Wen, Dayin Wang</i>	

<b>Process Management of Civil Aircraft Development Project Based on Workflow</b> .....	587
<i>Zeng Yuqin, Yan Guangrong, Zhao Gang</i>	
<b>JMF-Based Video Surveillance and Playback</b> .....	597
<i>Gao Xiuhui, Wang Wei, Zhao Gang</i>	
<b>New Aspects of Phoneme Synthesis Based on Chaotic Modeling</b> .....	605
<i>Marius Crisan</i>	
<b>Image Processing Scheme for Mobile Communication Equipment</b> .....	615
<i>Ye Liang</i>	
<b>Realization of a Multimedia Conferencing System Based on Campus Network</b> .....	623
<i>Wang Yun</i>	
<b>Study on the Wireless Access Communication of Line Condition Monitoring System of Smart Transmission Grid</b> .....	629
<i>Jinghong Guo, Hao Zhang, Yaxing Liu, Hongbin Liu, Ping Wang</i>	
<b>Forgery Attacks of a Threshold Signature Scheme with Fault-Tolerance</b> ...	637
<i>Yingda Yu, Chuanda Qi, Junjie He, Yuefen Chen</i>	
<b>Performance Assessment and Detection of Instant Messaging Software Based on the AHP&amp;LCA Model</b> .....	647
<i>ZongRen Liu, LeJun Chi</i>	
<b>Chinese Accents Identification with Modified MFCC</b> .....	659
<i>Qin Yan, Zhengjuan Zhou, Shan Li</i>	
<b>A 24GHz Microstrip Array Antenna with Low Side Lobe</b> .....	667
<i>Zhang Zhong-xiang, Chen Chang, Wu Xian-liang, Fu Ming-gang</i>	
<b>Resources Allocation of Multi-project in Railway Construction</b> .....	673
<i>Zhang Yunning, Jin Yao</i>	
<b>The Research on Development Risk of Green Construction Material Based on Fuzzy Comprehensive Evaluation</b> .....	679
<i>Zhang Yunning, Liu Gang</i>	
<b>Design and Development of Provincial Master Station of Metering Automation System of Guangdong Power Grid</b> .....	687
<i>Yang Jinfeng, Chen Ruimin, Xiao Yong, Dang Sanlei, Wei Fusheng</i>	
<b>The Research of Acceptance and Scheduling Scheme Based on the Overall Network QoS</b> .....	695
<i>Xu Zhanyang, Huang Li Ping</i>	

<b>Template-Based Image Composition on the iPhone</b> .....	705
<i>Qingjie Sun, Wenjuan Chen, Lei Liu</i>	
<b>Bag-of-Visual-Words Based Object Retrieval with E<sup>2</sup>LSH and Query Expansion</b> .....	713
<i>Zhao Yongwei, Li Bicheng, Gao Haolin</i>	
<b>Online Learning Algorithm of Direct Support Vector Machine for Regression Based on Matrix Operation</b> .....	727
<i>Junfei Li, Xiaomin Chen</i>	
<b>Cable Geometric Modeling Technology in Virtual Environment</b> .....	737
<i>Ma Long, Liu Pengyuan, Mi Shuangshan</i>	
<b>Communication between Vehicle and Nomadic Device Used for Assistance and Automation</b> .....	747
<i>Daniel Thielen, Christian Harms, Tobias Lorenz, Frank Köster</i>	
<b>Texture Synthesis Based on Wang Tiles</b> .....	759
<i>Yan Liu, GuangLin Cheng</i>	
<b>Research on the Control Strategy of the Driving System for a Kind of Hybrid Electric Vehicle</b> .....	765
<i>Lu Fang, Wang Deng-feng, Li Da-peng</i>	
<b>The Wireless Ranging System Based CSS Technology</b> .....	773
<i>Sun Hao, Wang Xin, Lv Changzhi, Gong Weimin</i>	
<b>Research on Communication Network of Smart Electricity Utilization for Residence Community</b> .....	781
<i>Zhang Xin-lun, Wang Zhong-jian, Bu Xian-de, Zhu Yu-jia</i>	
<b>Security Audit Based on Dynamic ACCESS CONTROL</b> .....	787
<i>Li Weiwei, Zhang Tao, Lin Weimin, Deng Song, Shi Jian, Wang Chen</i>	
<b>A Study on Information Systems Risk Management Based on Control Self-assessment</b> .....	795
<i>Gengshen Yu, Xiaohuan Zhu</i>	
<b>An Implementation of User Interaction Effect by Extending jQuery in Web Application</b> .....	801
<i>Seon Ung Lee, H-Young Moon</i>	
<b>An Intelligent Agent-Based Framework for Information Security Management</b> .....	807
<i>Chengzhi Jiang, Bo Zhang, Yong Yu, Xiaojian Zhang</i>	
<b>Effect of Atmospheric Turbulence Coherence Degradation on the Scintillation Index for a Partially Coherent Laser Beam</b> .....	815
<i>Ruike Yang, Xiaoning Kang, Xiaobo Zhang</i>	

<b>The Application of Mobile-Learning in Collaborative Problem-Based Learning Environments</b> .....	823
<i>Chien Yun Dai, Tzu-Wei Chen, Dar-Chin Rau</i>	
<b>Modeling Parallel MPI Programs in Petri Nets</b> .....	829
<i>Peng Zhang, Mei Qi</i>	
<b>The Design and Implementation of the IDS Catalogue Data Archive System</b> .....	837
<i>Li Jingang, Liu Shibin, Liu Wei</i>	
<b>Dem Generation with ALOS_PRISM Stereo Data for Zijinshan Diggings</b> ...	845
<i>Xiaomei Zhang, Guojin He, Chunxiang Chen</i>	
<b>A Novel Smart Home System</b> .....	853
<i>Ligeng Yu, Xueen Li, Weijuan Han</i>	
<b>Problem and Countermeasure on the Development of Mobile Electronic Commerce in China</b> .....	861
<i>Jiangjing Wang, Jiulei Jiang, Feng He</i>	
<b>Information Inquiry and Volume Analysis System on Urban Traffic Based on VB and MATLAB</b> .....	867
<i>Xiangcai Zhu, Yingkun Hou, Jian Xu, Yuncai Luan</i>	
<b>Multivariate Analysis on System Reliability and Security</b> .....	875
<i>Xiaoyan Wu, Zihui Che, Jing Jia</i>	
<b>Design and Application of a Network Platform for Electronic Contract Making</b> .....	883
<i>Yang Jianzheng, Xia Yunchao</i>	
<b>Implementation of the Building Decoration Design Demonstration System Based on Virtual Reality</b> .....	891
<i>Wu Feng, Wang Wei, Liu Xiuluo</i>	
<b>Expanded Study on Criticality in Schedule Plan of Project</b> .....	899
<i>Li Dayun, Guo Shu-hang</i>	
<b>WCF Applications in RFID-Based Factory Logistics Management System</b> .....	913
<i>Zhang Jianqi, Li Changzheng, Bao Fumin</i>	
<b>Research on Construction Method of Multimedia Semantic Model for Intangible Cultural Heritage</b> .....	923
<i>Guoxin Tan, Tinglei Hao, Shaoquan Liang, Zezhao Lin</i>	
<b>Improved Photonic Crystal 90° Bends for Millimeter—Wave Transmission</b> .....	931
<i>Q.Z. Xue, S.J. Wang</i>	

<b>Simulation of Algorithm of Binary Offset Carrier Signal</b> .....	935
<i>Xu Jia, Bi Mingxue</i>	
<b>The Performance Test Research for Application System of Internet of Things</b> .....	947
<i>Song Qizhu, Ma Aiwen, Peng Xiao, Wang Junfeng, Zhang Junchi</i>	
<b>The Selection of the Solar Battery Board Applied in New-Type Photovoltaic Generation Awnings</b> .....	953
<i>Xu-ping Chen, De-min Xiong</i>	
<b>Unbalanced Three-Phase Power Flow Calculation Based on Newton Method for Micro-Grid</b> .....	965
<i>Jiang Guixiu, Shu Jie, Wu Zhifeng, Zhang Xianyong</i>	
<b>Research of the Design of a Scene Simulation General Framework</b> .....	973
<i>Yao Linhai, Kang Fengju, Han Hong</i>	
<b>Research on the Prediction of Web Application System Aging Trend Oriented to User Access Intention</b> .....	983
<i>Jun Guo, Hao Huang, Bo Wang, Yunsheng Wang</i>	
<b>The Dynamic Control of the Process Parameters of MAG</b> .....	993
<i>Chuntian Li, Changhua Du, Huibin Xu, Yi Luo</i>	
<b>Analysis of Occupational Accidents during Construction of Buildings Using Classification and Regression Tree</b> .....	1003
<i>Chia-Wen Liao</i>	
<b>Weapons Program Based on an Integrated Management Model of the Hall</b> .....	1011
<i>Yongtao Yu, Ying Ding</i>	
<b>Opinion Search and Opinion Mining: A New Type of Information Technology Bringing about New Scientific and Talent Revolution</b> .....	1019
<i>Liu Yidong</i>	
<b>Author Index</b> .....	1025

# Range of Query Processing in Peer to Peer Networks

R. Saravanan<sup>1</sup> and P. Vivekananth<sup>2</sup>

<sup>1</sup> Lecturer, St. Joseph College of Engineering and Technology, Dar Es Salaam, Tanzania

<sup>2</sup> Lecturer, Botho College, Gaborone, Botswana

{rlsaravanan,vivek.jubilant}@gmail.com

**Abstract.** Peer to peer databases are becoming prevalent on the Internet for distribution and sharing of documents, applications, and other digital media. The problem of answering large-scale ad hoc analysis queries, such as aggregation queries, these databases poses unique challenges. Exact solutions can be time consuming and difficult to implement, given the distributed and dynamic nature of P2P databases. In this paper, we present novel sampling-based techniques for approximate answering of ad hoc aggregation queries in such databases. Computing a high-quality random sample of the database efficiently in the P2P environment is complicated due to several factors: the data is distributed across many peers, within each peer, the data is often highly correlated, and, moreover, even collecting a random sample of the peers is difficult to accomplish. To counter these problems, we have developed an adaptive two-phase sampling approach based on random walks of the P2P graph, as well as block-level sampling techniques. We present extensive experimental evaluations to demonstrate the feasibility of our proposed solution.

**Keywords:** Approximate query processing, aggregate query processing, computer networks, distributed database query processing, query cost estimation, ideal two phase algorithm.

## 1 Introduction

### 1.1 Peer-to-Peer Databases

The P2P network model is becoming the preferred medium for file sharing and distributing data over the Internet. A P2P network consists of numerous peer nodes that share data and resources with other peers on an equal basis. Unlike traditional client-server models, no central coordination exists in a P2P system; thus, there is no central point of failure. P2P networks are scalable, fault tolerant, dynamic and nodes can join and depart the network with ease. The most compelling applications on P2P systems to date have been file sharing and retrieval. Furthermore, researchers have been interested in extending sophisticated infrared techniques such as keyword search and relevance retrieval to P2P databases.

## 1.2 Aggregation Queries

The problem on P2P systems that is different from the typical search and retrieval applications. As P2P systems mature beyond file sharing applications and start getting deployed in increasingly sophisticated e-business and scientific environments, the vast amount of data within P2P databases poses a different challenge that has not been adequately researched so far, that is, how aggregation queries on such databases can be answered. Aggregation queries have the potential of finding applications in decision support, data analysis, and data mining. For example, millions of peers across the world may be cooperating on a grand experiment in astronomy, and astronomers may be interested in asking decision support queries that require the aggregation of vast amounts of data covering thousands of peers. Sensor networks can directly benefit from aggregation of traffic analysis data by offering a more efficient means of computing various network-based aggregates such as the average message size and maximum data throughput within the network, with minimal energy consumption and decreased response times. The problems are precise as follows: Consider a single table T that is distributed over a P2P system; that is, the peers store horizontal partitions of the table. An aggregation query such as the following may be introduced at any peer.

### Aggregation Query

```
SELECT Agg-Op(Col) FROM T WHERE election- condition
```

In the above query, the Agg-Op may be any aggregation operator such as SUM, COUNT, AVG, and so on, Col may be any numeric measure column of T or even an expression involving multiple columns, and the selection condition decides which tuples should be involved in the aggregation. Although our main focus is on the above standard SQL aggregation operators, we also briefly discuss other interesting statistical estimators such as medians, quantiles, histograms, and distinct values. Although aggregation queries have been heavily investigated in traditional databases, it is not clear that these techniques will easily adapt to the P2P domain. For example, decision support techniques such as online analytical processing commonly employ materialized views; however, the distribution and management of such views appear difficult in such a dynamic and decentralized domain. In contrast, the alternative of answering aggregation queries at runtime “from scratch” by crawling and scanning the entire P2P repository is prohibitively slow.

## 1.3 Approximate Query Processing

Fortunately, it has been observed that in most typical data analysis and data mining applications, timeliness and interactivity are more important considerations than accuracy; thus, data analysts are often willing to overlook small inaccuracies in the answer, provided that the answer can be obtained fast enough. This observation has been the primary driving force behind the recent development of AQP techniques for aggregation queries in traditional databases and decision support systems[1] [3], [7], [8], [9] [10]. Numerous AQP techniques have been developed: The most popular ones are based on random sampling, where a small random sample of the rows of the



database is drawn, the query is executed on this small sample, and the results are extrapolated to the whole database. In addition to simplicity of implementation, random sampling has the compelling advantage that, in addition to an estimate of the aggregate, one can also provide confidence intervals of the error, with high probability. Broadly, two types of sampling-based approaches have been investigated: first the pre computed samples, where a random sample is pre computed by scanning the database and the same sample is reused for several queries and second online samples, where the sample is drawn “on the fly” upon encountering a query.

### **Nomenclature**

P2P	– Peer to Peer
AQP	– Approximate query processing
BFS	– Breadth first search
DFS	– Depth first search
IR	– Infrared
OLAP	– Online analytical processing
IPp	– Processor address
PORTP	– Port number
TCP	– Transmission control protocol
IP	– Internet protocol

## **2 P2P Sampling Process Using Skew Random Walk**

It describes the framework of the approach. Essentially trying to pick true uniform random samples of the tuples, as such samples are likely to be extremely impractical to obtain. Instead, of consider an approach where willing to work with skewed samples, provided that accurately estimate the skew during the sampling process. To get the accuracy in the query answer desired by the user, skewed samples can be larger than the size of a corresponding uniform random sample that delivers the same accuracy; however, the processes samples are much more cost efficient to generate. Certain aspects of the P2P graph are known to all peers, such as the average degree of the nodes, a good estimate of the number of peers in the system, certain topological characteristics of the graph structure, and so on. Estimating these parameters via preprocessing are interesting problems in their own right; parameters are relatively slow to change and thus do not have to be estimated at query time: It is the data contents of peers that changes more rapidly; hence, the random sampling process that picks a representative sample of tuples has to be done at runtime. The approach has two major phases. In the first phase, initiate a fixed-length random walk from the query node. This random walk should be long enough to ensure that the visited peers represent a close sample from the underlying stationary distribution. Then retrieve certain information from the visited peers, such as the number of tuples, the aggregate of tuples such as SUM, COUNT, AVG, and etc; that satisfy the selection condition, and send this information back to the query node. This information is then analyzed at the query node to determine the skewed nature of the data that is distributed across the network, such as the variance of the aggregates of the data at peers, the amount of

correlation between tuples that exists within the same peers, the variance in the degrees of individual nodes in the P2P graph recall that the degree has a bearing on the probability that a node will be sampled by the random walk and so on. Once this data has been analyzed at the query node, an estimation is made on how much more samples are required so that the original query can be optimally answered within the desired accuracy, with high probability. For example, the first phase may recommend that the best way to answer this query is to visit  $m'$  more peers and, from each peer, randomly sample  $t$  tuples. The first phase is not overly driven by heuristics. Instead, it is based on underlying theoretical principles such as the theory of random walks [6] as well as statistical techniques such as cluster sampling, block-level sampling, and cross validation. The second phase is then straightforward: A random walk is reinitiated, and tuples are collected according to the recommendations made by the first phase. Effectively, the first phase is used to “sniff” the network and determine an optimal-cost “query plan,” which is then implemented in the second phase. For certain aggregates such as COUNT and SUM, further optimizations may be achieved by pushing the selections and aggregations to the peers; that is, the local aggregates instead of raw samples are returned to the query node, which are then composed into a final answer.

### 3 An Application of P2P Protocol Using Gnutella

Assume an unstructured P2P network represented as a graph  $G = (P, E)$  with a vertex set  $P = \{p_1, p_2, p_3 \dots p_m\}$  and an edge set  $E$ . The vertices in  $P$  represent the peers in the network, and the edges in  $E$  represent the connections between the vertices in  $P$ . Each peer  $p$  is identified by the processor's IP address and a port number (IP $_p$  and port $_p$ ). The peer  $p$  is also characterized by the capabilities of the processor on which it is located, including its CPU speed  $pcpu$ , memory and width  $pmem$ , and disk space  $pdisk$ . The node also has a limited amount of bandwidth to the network, noted by  $pband$ . In unstructured P2P networks, a node becomes a member of the network by establishing a connection with at least one peer currently in the network. Each node maintains a small number of connections with its peers: The number of connections is typically limited by the resources at the peer. The number of connections in a peer is maintained by  $pconn$ . The peers in the network use the Gnutella P2P protocol to communicate. The Gnutella P2P protocol supports four message types Ping, Pong, Query, and Query\_Hit, of which the Ping and Pong messages are used to establish connections with other peers, and the Query and Query\_Hit messages are used to search in the P2P network. Gnutella, however, uses a naive Breadth First Search technique in which queries are propagated to all the peers in the network and thus consumes excessive network and processing resources and results in poor performance. The approach, on the other hand, uses a probabilistic search algorithm based on random walks. The key idea is that each node forwards a query message, called walker, randomly to one of its adjacent peers. This technique is shown to improve the search efficiency and reduce unnecessary traffic in the P2P network.

## 4 Measuring of Query Cost Using Latecy Algorithm

The execution cost of a query in P2P databases is more complicated than equivalent cost measures in traditional databases. The primary cost measure is considering latency, which is the time that it takes to propagate the query across multiple peers and receive replies at the query node. In our algorithm, latency can be approximated by the number of peers that participate in the random walk. This measure is appropriate for our algorithm because it performs a single random walk starting from the query node. Thus, latency becomes proportional to the total number of visited peers in the random walk.

The aggregation operator can be pushed to each visited peer. Once a peer is visited by the algorithm, the peer can be instructed to simply execute the original query on its local data and send only the aggregate and the degree of the node back to the query node, from which the query node can reconstruct the overall answer. Moreover, this information can be sent directly without necessitating any intermediate hops, as the visited peer knows the IP address of the query node from which the query originated. This is reasonable, considering that the IP address can be pushed to visited peers along with the aggregation operator and the P2P networks such as Kazaa run on top of a TCP/IP layer, making it feasible to make direct connections with peers. Thus, the bandwidth requirement of such an approach is uniformly very small for all visited peers: They are not required to send more voluminous raw data back to the query node. In approximating latency by the number of peers participating in the random walk make the implicit assumption that the overhead of visiting peers dominates the costs of local computations such as execution of the original query on the local database. This is, of course, true if the local databases are fairly small. To ensure that the local computations remain small even if local databases are large, our approach in such cases is to execute the aggregation query only on a small fixed-sized random sample of the local data (that is a subsample from the peer), scale the result to the entire local database, and send the scaled aggregate back to the query node. This way ensures the local computations are uniformly small across all visited peers.

## 5 The References Section

The theorem says that the expected value of the square of the CVError is two times the expected value of the square of the actual error. The CVError can be estimated in the first phase by the following procedure. Randomly divide the  $m$  samples into halves and compute the CVError (for sample size  $m=2$ ). Then determine  $C$  by fitting this computed error and the sample size  $m=2$  into the equation in Theorem.

$$E[\text{CVError}^2] = 2E[(y''-y)^2]$$

Proof

$$E[\text{CVError}^2] = E[(y1''-y2'')^2]$$

$$\begin{aligned}
&= E[(y_1''-y)^2] + E[(y_2''-y)^2] \\
&= 2E[(y''-y)^2]
\end{aligned}$$

To get a somewhat more robust estimation for  $C$ , we can repeat the random having of the sample collected in the first phase several times and take the average value of  $C$ . Note the CVError is larger than the true error, the value of  $C$  is conservatively overestimated. Once  $C$  is determined that is, the “badness” of the clustering of data in the peers, then determine the right number of peers to sample in the second phase  $m'$  to achieve the desired accuracy.

## 6 Ideal Two Phase Algorithm

In this section, contains the details of the two-phase algorithm for approximate answering of aggregate queries. For illustration focus on approximating COUNT queries it can be easily extended to SUM, AVERAGE and MEDIAN queries. In the first phase is broken up into the following main components. First A random walk on the P2P network, attempting to avoid skewing, due to graph clustering and vertices of high degree. Walk skips  $j$  nodes between each selection to reduce the dependency between consecutive selected peers. As the jump size increases, our method increases overall bandwidth requirements within the database, but for most cases, small jump sizes suffice for obtaining random samples. Second compute aggregates of the data at the peers and send these back to the query node. This was primarily done to keep the previous discussion simple. In reality, the local databases at some peers can be quite large, and aggregating them in their entirety may not be negligible as compared to the overhead of visiting the peer. In other words, the simplistic cost model of only counting the number of visited peers is inappropriate. In such cases, it is preferable to randomly subsample a small portion of the local database and apply the aggregation only to this subsample. Thus, the ideal approach for this problem is to develop a cost model that takes into account cost of visiting peers, as well as local processing costs. Moreover, for such cost models, an ideal two-phase algorithm should determine various parameters in the first phase, such as how many peers should be visited in the second phase and how many tuples should be subsampled from each visited peer. In this paper, the simple approach in to fix a constant  $t$  determined at preprocessing time via experiments such that if a peer has at most  $t$  tuples, then its database is aggregated in its entirety, whereas if the peer has more than  $t$  tuples, then  $t$  tuples are randomly selected and aggregated. Subsampling can be more efficient than scanning the entire local database, for example, by blocklevel sampling, in which only a small number of disk blocks are retrieved. If the data in the disk blocks are highly correlated, then it will simply mean that the number of peers to be visited will increase, as determined by our crossvalidation approach at query time. Third, we estimate the CVError of the collected sample and use that to estimate the additional number of peers that need to

be visited in the second phase. For improving robustness, steps in the cross-validation procedure can be repeated a few times, as well as the average squared CVError computed. Once the first phase has completed, the second phase is then straightforward. This paper simply initiates a second random walk based on the recommendations of the first phase and computes the final aggregate.

### 6.1 Estimation of Aggregate Queries Using Sum and Average

Although the algorithm has been presented for COUNT queries, it can be easily extended to other aggregates such as the SUM and AVERAGE by modifying the  $y(\text{Curr})$  value specified of the algorithm.

### 6.2 Estimation of Aggregate Queries Using Median

For more complex aggregates such as estimation of medians and distinct values, more sophisticated algorithms are required. In addition to computing COUNT, SUM, and AVERAGE aggregates, can also efficiently estimate more difficult aggregates such as the MEDIAN. Propose an algorithm for computing the MEDIAN Figure 4. in a distributed fashion based upon comparing the rank distances of medians of individual peers. Our algorithm for computing the MEDIAN is given as follows:

1. Select  $m$  peers at random by using random walk
2. Each peer  $s_j$  computes its median  $med_j$  and sends it to the query node along with  $prob(s_j)$
3. The query node randomly partitions the  $m$  medians into two groups of  $m/2$  medians : group1 and group2.
4. Let  $med_{g1}$  be the weighted median of group1, that is such that the following is minimized

$$abs \left( \sum_{\substack{med_j \in group1, \\ med_j \angle med_{g1}}} \frac{1}{prob(s_j)} - \sum_{\substack{med_j \in group1, \\ med_j \geq med_{g1}}} \frac{1}{prob(s_j)} \right)$$

5. Find the error between the median of group2 (say,  $med_{g2}$ ) and the weighted rank of  $med_{g1}$  in group2. That is, let

$$abs \frac{\left( \sum_{\substack{med_j \in group2, \\ med_j \angle med_{g1}}} \frac{1}{prob(s_j)} - \sum_{\substack{med_j \in group2, \\ med_j \geq med_{g2}}} \frac{1}{prob(s_j)} \right)}{m/2}$$

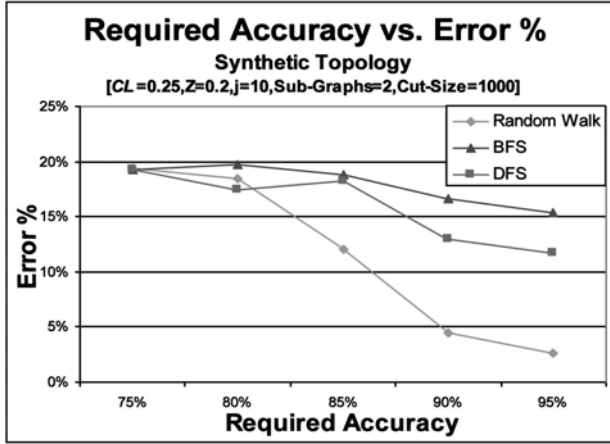


Fig. 1. Accuracy comparison of random walk with BFS & DFS

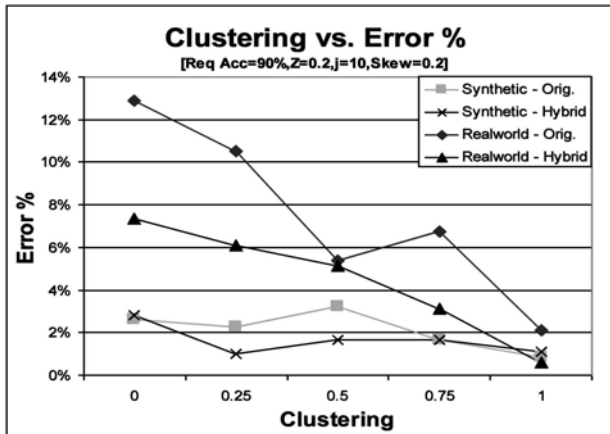


Fig. 2. Effects of clustering on the error percentage for the COUNT technique

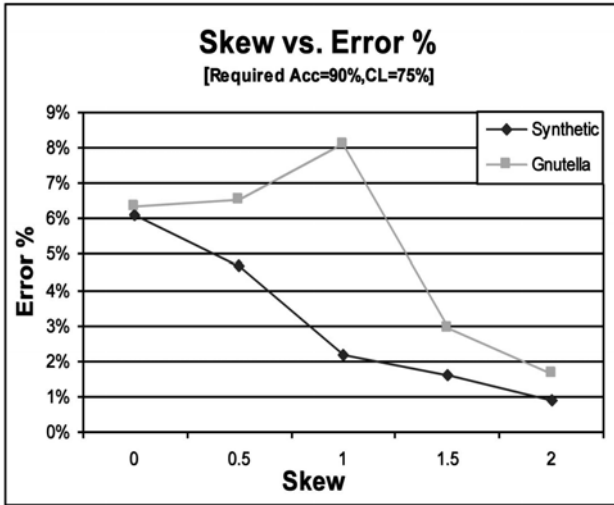


Fig. 3. Effects of clustering on the error percentage for the SUM technique

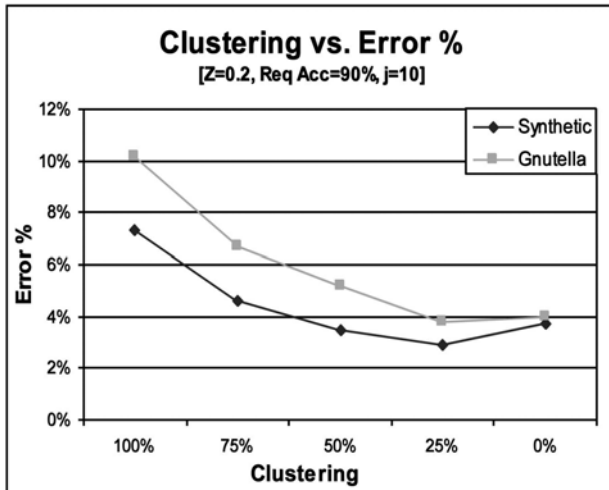


Fig. 4. Effects of clustering on the error percentage for the MEDIAN technique

## 7 Conclusions

The proposed adaptive sampling-based techniques are for the approximate answering of ad hoc aggregation queries in P2P databases. This approach requires a minimal number of messages sent over the network and provides tunable parameters to

maximize performance for various network topologies. The approach provides a powerful technique for approximating aggregates of various topologies and data clustering but comes with limitations based upon a given topologies structure and connectivity. For topologies with very distinct clusters of peers small cut size, it becomes increasingly difficult to accurately obtain random samples due to the inability of random-walk process to quickly reach all clusters. This can be resolved by increasing the jump size, allowing a larger number of peers to be considered and increasing the allowed mixing by the proposed hybrid approach. By varying a few parameters, our algorithm successfully computes aggregates within a given required accuracy. The presented extensive experimental evaluations are to demonstrate the feasibility of our solutions for both synthetic and real-world topologies.

**Acknowledgments.** The author gratefully acknowledges the authorities of St. Joseph college of Engineering and Technology, Botho colleges for the facilities offered to carry out this work.

## References

1. Acharya, S., Gibbons, P.B., Poosala, V.: Aqua: A Fast Decision Support System Using Approximate Query Answers. In: Proc. 25th Int'l Conf. Very Large a data Bases, VLDB 1999 (1999)
2. Adamic, L., Lukose, R., Puniyani, A., Huberman, B.: Search in Power-Law Networks. *Physical Rev. E* (2001)
3. Babcock, B., Chaudhuri, S., Das, G.: Dynamic Sample Selection for Approximate Query Processing. In: Proc. 22nd ACM SIGMOD Int'l Conf. Management of Data (SIGMOD 2003), pp. 539–550 (2003)
4. Bharambe, A.R., Agrawal, M., Seshan, S.: Mercury: Supporting Scalable Multi-Attribute Range Queries. In: Proc. ACM Ann. Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm., SIGCOMM 2004 (2004)
5. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Analysis and Optimization of Randomized Gossip Algorithms. In: Proc. 43rd IEEE Conf. Decision and Control, CDC 2004 (2004)
6. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Gossip and Mixing Times of Random Walks on Random Graphs. In: Proc. IEEE INFOCOM 2005 (2005)
7. Charikar, M., Chaudhuri, S., Motwani, R., Narasayya, V.: Towards Estimation Error Guarantees for Distinct Values. In: Proc. 19th ACM Symp. Principles of Database Systems, PODS 2000 (2000)
8. Chaudhuri, S., Das, G., Datar, M., Motwani, R., Narasayya, V.: Overcoming Limitations of Sampling for Aggregation Queries. In: Proc. 17th IEEE Int'l Conf. Data Eng. (ICDE 2001), pp. 534–542 (2001)
9. Chaudhuri, S., Motwani, R., Narasayya, V.: Random Sampling for Histogram Construction: How Much Is Enough. In: Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD 1998), pp. 436–447 (1998)
10. Chaudhuri, S., Das, G., Narasayya, V.: A Robust Optimization- Based Approach for Approximate Answering of Aggregate Queries. In: Proc. 20th ACM SIGMOD Int'l Conf. Management of Data, SIGMOD 2001 (2001)



# A Scene Detection Based Scheme for Supporting VCR Functions

Shu-bin Xu

The 54th Research Institute of CETC, Shijiazhuang, Hebei, 050081, P.R. China

**Abstract.** In a video streaming system, the VCR operations such as fast forward, fast backward etc, need more network bandwidth than the normal playback due to the transmitting of redundant frames. In this paper, a novel scene-detection scheme was proposed, and introduced to the implement method of fast forward operation. The scheme allocates bandwidth according to each scene's length, and select frames in each scene to be displayed dynamically. Experimental results show that the scheme can implement the fast forward operation at a bandwidth which was allocated for the normal playback situation, meanwhile keep the frame rate in the range of 22 to 27 and the mean average activity greater than 95 percent until the speed factor 24.

**Keywords:** VOD, VCR operation, Scene detection, Fast backward.

## 1 Introduction

With the development of broadband networking and computing technology, video applications such as interactive TV, video on demand (VOD), video conference play an important role in modern multimedia communications. With the profiting of video compression standard (MPEG-4, H.264), broadband access networking, digital set-top box(STB),VOD which is one of the prominent service of streaming applications becomes popular. However its high bandwidth requirement[1] for browsing the videos makes it still a challenging to support VCR operations[2].

Supporting interactive and VCR (Video Cassette Recorder) functionalities such as fast-forward (FF), pause, rewind etc is the development trend of modern video on demand systems. The traditional implemental methods of k-times FF operation is displayed one frame every other k frames to the user. As a result, in the FF operation, extra m frames was transmitted in order to decode the P or B frames. This special requirement leads to a much higher network and decoder complexity[3]. There are many researches in how to reduce the bandwidth requirement of the VCR[4,5], but the focus is mainly based of the I-B-P frame's relationship and the size, the understanding of the frame sequence displayed to the user during the FF operation was not one of the important factors in their schemes.

The research in this paper, how to implement the VCR operation with a fixed bandwidth which is usually allocated for the normal playback of the video was considered. and video scene detection was introduced into video streaming VCR operations. Since scene is helpful for video content understanding, in this paper, a novel scene-detection based scheme for VCR operations (called SDSV) was

proposed. The SDSV scheme can bring such advantages compare to traditional VCR methods: 1) the users can obtain more information when they select fast forward or fast backward; 2) it can significantly reduces the network bandwidth requirement.

## 2 Model of VCR

The VCR operation especially fast forward(FF) and fast backward(FB) in video streaming applications usually produce much higher network bandwidth requirement compared with normal play. In this paper, the FF operations was focused, and the FB is similar with FF.

The system prototype in this paper is described as follows: clients request a certain video file for the video streaming server, if the network bandwidth is enough for the client, then the server allocate a fixed bandwidth for it, and the bandwidth keep unchanged despite of what operations the user select, such as FF, FB, PAUSE, REWIND. The VCR implementation method must meet the bandwidth constraint condition.

**Definition:** Given the start number of the FF operation is  $F_s$ , end number is  $F_e$ , the speed factor is  $k$  and frame rate is  $F$ , then the duration time of this clip is  $T_n = (F_s - F_e)/F$ , and the user expects to view the video clip in  $T_e = T_n/k$  time length, meanwhile, have a approximately impression of the clip. Let  $F_i$  be a frame needed to be displayed to the user, then its display time is  $(F_i - F_s)/k$ .

Let  $S = \{S_1, S_2, S_3, S_4 \dots S_M\}$ , where  $S_i$  is the  $i$ th frame size of the video, and  $M$  is the total frame number.  $D = \{D_1, D_2, D_3, \dots D_{d(k)}\}$ , where  $D$  is the displayed frame sequence to the user under speed factor  $k$ , and  $d(k)$  is the total number of displayed frames.

$\delta(t_i)$  denotes the duration time of frame  $D_i$ , note that the  $\delta(t_i)$  is not must be same with  $1/F$  in the FF situation. Let  $G = \{G_1, G_2, G_3, \dots G_{d(k)}\}$ , where  $G_i$  is the total size the server send to the client to display the frame  $D_i$ ,  $D_i$  is not always equal to  $S(D_i)$ , because if  $D_i$  is a P or B frame, the server should send its reference frames. Given the bandwidth reserved for the client is  $B$  when the user requested for a video the first time. Our aim is to design the sequence  $D$  and  $G$  to satisfy the equations as follows:

$$\frac{G_i}{\Delta t_i} \leq B \quad (1)$$

$$\sum_{i=1}^{d(k)} \Delta t_i = \frac{F_e - F_s}{k \times F} \quad (2)$$

Eq.1 is the bandwidth constraint, and Eq.2 is the time constraint to satisfy the user's  $k$  times FF experience.

## 3 Scene-Detection Based VCR Implemental Scheme

The video structure can be represented based on the video-scene-shot-frame hierarchy. Video is a sequence of frames which the content is changed along with the

time, and frame is its minimum unit. Shot is the basic unit of video, it's the camera's consecutive shooting. Scene is a cluster of a series of shots, which have a similar plot, describing a same event in a different view point. So scene is the minimum unit of a video which can be understand, it is the semantic unit of the video.

It makes sense to consider why a user choosing the FF operations before we implement it. Thus, we categorize the purposes of a user choosing FF as follows: a) Find the interesting point (IP), then start to view the video, meanwhile, get a overview of the contend between the start point and the IP. b) Locate quickly to the point that he/she stopped to play last time for some reasons, but at the same time, get a previous overview of the storyline from the beginning to the point. c) Browse quickly to determine whether this video is his/her type.

From the previous analysis, we can conclude that the quality of user's perception (QoP) will be better if they understand the scene even partially during the FF operations. However, most of the researchers developed their algorithms based on how to reduce the bandwidth that the FF operation occupied, and the scene was not one of these factors they considered about. Usually, these algorithms introduce some intelligent frame selection methods in a GOP to meet the bandwidth constraint conditions such as Eq.2. In this paper, we believe that if put the scene factor and bandwidth factor together, the user will get a better perception during the FF operation. Thus, we presented the SDSV scheme, which proposed a scene based FF operation scheme.

The basic idea of SDSV is assigning each scene's playing time during FF operation according to the scene's length, meanwhile, based on the bandwidth constraint, deciding which frames need to be displayed.

Let  $F = \{F_{1,y1}, F_{2,y2}, F_{3,y3}, F_{4,y4}, \dots, F_{N,yN}\}$ , where  $F_{i,yi}$  is the  $i$ th scene of the video,  $y_i$  is the total frame number that the  $i$ th scene contains, and  $N$  is the total scenes number. So  $F_{i,yi} = \{j, j+1, j+2, \dots, j+y_i\}$ , where  $j$  is the first frame number in the scene, and  $S_j$  is the frame size. For a better QoP from the above discussion, we defined another metric for the FF operation: the scene's average activity ( $S_a$ ), and  $S_a = S_d/N$ , where  $S_d$  means the number of displayed scenes to the user, and  $S_d = \sum_{i=1}^N \mathcal{E}_i$ ,  $\mathcal{E}_i = 1$  if there exist at least one frame that displayed to the user in the  $i$ th scene, otherwise  $\mathcal{E}_i = 0$ . Obviously,  $S_a \in [0, 1]$ , in the normal playback situation, every scene is sent to the user, thus  $S_a = 1$ , and  $S_a = 0$  if there is no frame been sent. From the discussion we derive that during the FF operation the closer  $S_a$  to 1 the more scenes displayed to the user and the better QoP we get. Given that the FF operation started at the first scene of the video, and ended until the last scene. Based on the definition of the FF operation and the assumption that we should display as much scenes as possible to the user in order to achieve a better QoP, it is acceptable that first we try to display every scene if it can meet the bandwidth constraint in Eq.1. We can calculate the  $i$ th scene's playback time during the FF operation as in the following:

$$T_k^i = \frac{y^i}{F \times k} \quad (3)$$

Where  $k$  is the speed factor, and  $F$  is the frame rate. The total bytes this scene can be played is as in the following:

$$G_i' = B \times T_k^i \tag{4}$$

$G_i'$  is the max size that the server can send to the client. Usually, a scene contains several GOPs, and a GOP contains several frames(at least one I-frame). After get  $G_i'$ , the SDSV scheme attempt to choose the displaying frames in these GOPs that the  $i$ th scene contains under the Eq.2.

To decode a P-frame, the preceding I-frame of it is always needed, so the scheme tries to allocate  $G_i'$  for the I-frames in the scene firstly. Let  $K_i$  denote the overall size of I-frames in the  $i$ th scene, and  $K_i^f$  denote the minimum size of the I-frames in the  $i$ th scene, then there are 3 conditions between  $G_i'$ , and  $K_i^f$ :

a)  $K_i^f \leq G_i' \leq K_i$

It is means that in the  $i$ th scene there is not enough bandwidth to send all the I-frames. Because the I-frames have the highest priority, we only send some of these I frames and don't consider other frames any more. Note that in this situation, this scene is considered being displayed and  $S_d$  should be augmented by 1.

b)  $K_i^f > G_i'$

In this situation, even only sending the minimum I-frame of this scene is still difficult to meet the bandwidth constraint. Therefore, this scene will not be displayed. For the purpose of obtaining the max  $S_a$ , the bandwidth allocated to it should be reassigned to its next scene. That is  $G_{i+1}' = G_{i+1}' + G_i'$ . Note that  $S_d$  should no be augmented.

c)  $G_i' > K$

Beside sending all the I-frames in this scene, there are still remaining some bytes for sending the P frames. In order to display as much as more frames to the user, the rule that the SDSV choose the P frames is based on the distance between the P-frame and its preceding I-frame. The shorter the distance is, the higher priority that being displayed the P-frame has. This rule is being described in detail as in Fig.1.

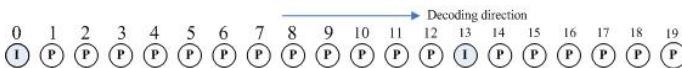


Fig. 1. P-Frame selection algorithm

In the Fig. 1, the scene contained two GOPs, and each GOP with one I-frame and twelve P-frames. First we choose the frame 0 13, then 0 1 13, if there still enough bytes then 0 1 13 14, then 0 1 2 13 14, the selection is carrying out until the bytes being consumed completely. Note that in this situation, this scene is considered being displayed and  $S_d$  should be augmented by 1.

## 4 Experimental Results

**Experimental Environment:** The experiment is based on a real-life MPEG-IV video trace, Star Wars IV, 30 minutes long. The mean bit rate of it is 3.38Mb/s, and the trace is played back at a rate of  $F=30$  frames/s in normal situation. The total frames of this trace is 53953, 3373 of them are I-frames, 10116 are P-frames and others are B-frames. We assume that when the user first request this trace, a bandwidth of 3.4Mb/s is assigned to this user and remain un-change regardless the mode that the user choose (normal play or k times FF) until the end of the trace.

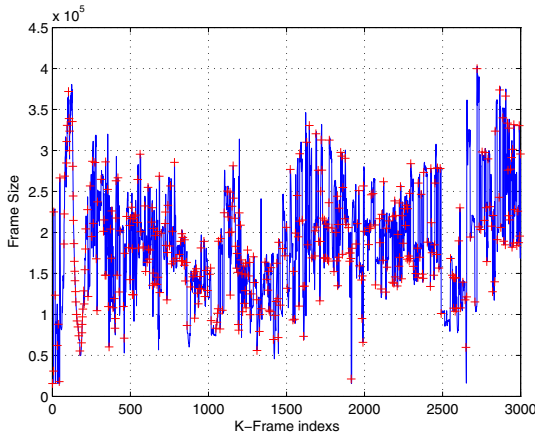


Fig. 2. Scene-detection results with ATDS

**Scene-change Detection:** The first step of our experiment is mark all the scenes of this trace. Researchers have done many works on this area[6]. In this paper, a scene detection scheme which called ATDS(Adaptive Threshold Determination Scheme for Scene Detection) was applied to identify the scenes in our trace. The ATDS was first proposed in [6], this scheme can dynamically identify the scene-change point through the size of I-frames, and the precision of it is about 90 percent. The scene-detection results were showed in Fig.2. The total number of scenes is 421 and all of them are marked with '+'. Note that the X-axis is frame index of the I-frames excluding P and B frames, because of the feature of ATDS.

**FF Operation Experimental Results:** Once the scene change points of the trace have been figured out, we can apply our SDSV scheme in the trace. We mainly analyzed the bandwidth requirement, frame rate and scene activity with different factor of the FF operation.

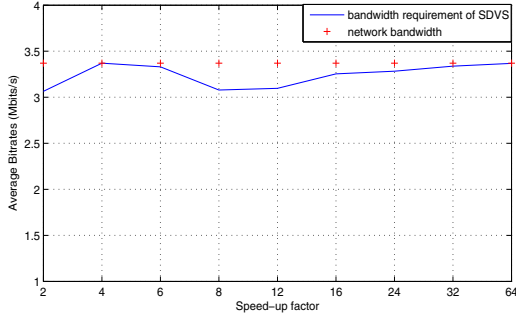


Fig. 3. Bandwidth requirement of SDSV

In the first step, the bandwidth requirement of the SDSV scheme with different speed factors was analyzed. The result was shown in Fig.3. The result show that the bandwidth that the SDSV needed was never exceeded the network bandwidth allocated to the user.

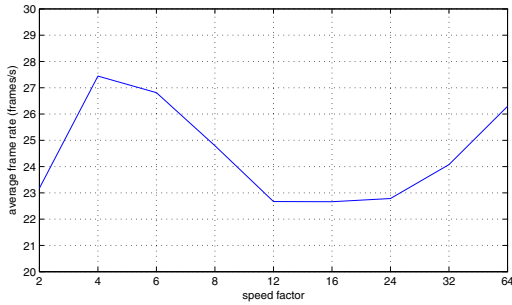


Fig. 4. Average frame rate of SDSV

In the second step, the average frame rate of our SDSV scheme with different speed factors was analyzed. In the normal playback situation, the frame rate is 30 frames/s, and in the FF operation the frame rate is varies in the range of 22 to 28, as shown in Fig.4. In fact, we believe that the decrease of the frame rate is very small, and it will not affect the user’s experience during the FF operation.

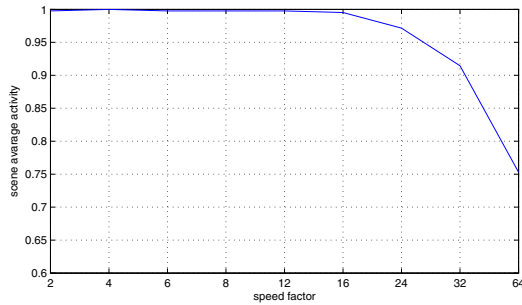


Fig. 5. Average scene-activity of SDSV

In the third step, the average scene activity of SDSV scheme with different speed factors was analyzed. If possible we should display every scene to the user for not loss any plot of the trace. From the results shown in Fig.5, one can conclude that most of the scenes are displayed to the user until speed factor 16. At the speed factor of 24, 32 and 64 the average scene activity decreased rapidly. This is because of there exists some rather short scenes and the speed factor is too fast for them and there is no bandwidth to display them.

## 5 Summary

In this paper, the difficulty of support VCR functionalities was reviewed. Based on the analysis of the VCR operation, a simple VCR function model was presented. According to the VCR model and the reason why the user choose a VCR operation we presented a novel scheme called SDSV. The principles of SDSV is assign each scene a total size that it can send based on the speed factor and the network bandwidth constraint, then select the frames to be displayed in each scene. Experiment result show that the scheme is effectively, it can support the VCR operation in the bandwidth constraint width a very high frame rate and scene activity.

Our ongoing researches will focus on the area of evaluation method of the user's experience in the situation of VCR operation.

**Acknowledgments.** This work is supported in part by Nation Grand Special Science and Technology Project of China under Grant No. 2010ZX03006-002-02.

## References

1. Li, V., Liao, W., Qiu, X., Wong, E.: Performance model of interactive video-on-demand systems. *IEEE Journal on Selected Areas in Communications* 14(6), 1099–1109 (1996)
2. Wu, D., Hou, Y., Zhang, Y.: Transporting real-time video over the internet: challenges and approaches. *Proceedings of the IEEE* 88(12), 1855–1877 (2000)
3. Lu, K., Li, Y., Qiao, C.: On guaranteed VoD services in next generation optical access networks. *IEEE Journal on Selected Areas in Communications* 28(6), 875–887 (2010)
4. Chen, M., Kandlur, D.: Downloading and stream conversion: supporting Interactive playout of videos in a client station. In: *IEEE 2nd Conf. on Multimedia Computing and Systems*, pp. 73–80 (1995)
5. Zhu, M., Li, J., Zhou, J.: VCR functionalities with adaptive rate adjustment in streaming media server. In: *IEEE 9th International Conference on Computer and Information Technology*, pp. 150–155 (2009)
6. Li, H., Liu, G., Zhang, Z., Li, Y.: Adaptive scene-detection algorithm for VBR video stream. *IEEE Transactions on Multimedia* 6(4), 624–633 (2004)

# Constructing Digital Campus Using Campus Smart Card System

Fang Wang and Zhuo-sheng Jia

Information Center of BeiJing Jiaotong University, BeiJing 100044, China  
brandy\_1977@hotmail.com, jia@bjtu.edu.cn

**Abstract.** Campus smart card system is basic supporting part in campus information constructions. It is one of the projects which can best reflect the effectiveness of campus information construction. It also plays a great role in improving school management and service. Campus smart card system is an integrated system of consumption and authentication. The systems collect a wealth of data. That everybody in the campus has a smart card extends the usefulness of smart card system. This paper describes the secondary developments and data mining based on campus smart card system in Beijing Jiaotong University. It lays emphasis on the design, function and implementation of Campus Personnel Data Center, Demission Procedure System, and Students Conducts Analyzing System.

**Keywords:** Digital campus, Campus Smart Card, Campus Personnel Data Center, Demission Procedure System, Student Conducts Analyzing System.

## 1 Introduction

Nowadays many colleges and universities are constructing campus smart card systems which are integrated systems of consumption and authentication, including several subsystems as catering, educational administration, finance, book management, medical care and access control.

Take our university as an example; the construction of smart card system is divided into two stages. In the first stage, the system builds 17 subsystems, including smart card center management system, catering management system, voice self-service platform and on-line service platform, internet fee transferring system, dormitory access control system, gate access control system, book borrowing system, new students management system (divided by student number and class), fingerprint registration system, computer room management system, dormitory management system, digital system for welcoming new students, conference attendance subsystem, examination supervision system, medical care subsystem, charging system of small amount, data synchronization system; and the second stage includes bank-university transferring system, medical records management system, library lockers system and vehicles management system, etc.

Such a large number of systems have collected plenty of data. We have made some attempts to explore how to make use of these data and how to analyze them.



## 2 Building Campus Personnel Center on the Basis of Campus Smart Card System

Personnel include different kinds. Take our university as an example; it comprises teachers, students and temporary workers. Teachers include teaching staff, retired staff, external employed staff, exchange scholars, post-docs, etc.; students include undergraduate students, graduate students, students of distance learning, students upgraded from associate-graduate and students of training classes. At present, teachers are in the charge of personnel department, external employed staffs are in the charge of each secondary units and students are in the charge of different educational administration departments, graduate schools, further education schools and training classes organized by each school. No one department alone administrates all the personnel. So it is extremely hard to conduct statistics of the whole personnel. Since the adopting of smart card, all personnel need to apply for smart card in the smart card center, because it relates to catering, bathing, hot water supplying, book borrowing, and registration. It is nearly impossible to live without smart card on campus. Thus, we came up with the idea of building personnel center on the basis of smart card.

### The Sources of Data

The data sources of our students are collected through:

**Table 1.** The data sources of our students

Category	Sources	Way of entering the system
undergraduates	Enrollment Department	On-line enrollment date enter the smart card system in the form of excel and are divided into different classes with student number. The data are processed in the smart card system.
postgraduate students	Graduate School	Personal information with student number and pictures are introduced into the system.
students of adult education	Further Education School	Personal information enters the system in excel, and pictures are taken at the spot.
students upgraded from associate-graduate	Educational Administration Department	Personal information enters the system in excel, and pictures are taken at the spot.
foreign students	Foreign Office	Personal information enters the system in excel, and pictures are taken at the spot.

The data sources of our teachers are collected through:

**Table 2.** The data sources of our teachers

Category	Sources	Way of entering the system
teachers	Labor Division of Personnel Department	The data are input into the system directly, and new teachers input the data with their identity card.
external employed teachers	External Employed Personnel Administration Office of Personnel Department	According to the contract affirmed by personnel department, external employed teachers are numbered and enter the smart card system.
retired teachers	Retirement Department	The data are input into the system directly, and the data are renewed when teachers go through the retirement formalities in the smart card center.
Post-docs	Teacher Qualification Division of Personnel Department	Personal information is input into the system in excel.
Exchange scholars	Teacher Qualification Division of Personnel Department	Personal information is input into the system in excel.

The data sources of our temporary workers are collected through:

**Table 3.** The data sources of our temporary workers

Category	Sources	Way of entering the system
students of all training classes	each school and department	Personal information is input into the system in excel.
Nonnative personnel	Canteen	The smart cards are distributed to the canteen, and the personal information is renewed when the cards are granted.

#### Change of Data

As for the change of personnel data, we will notify the related departments. Since there is not a general administration department, our university sets up a series of regulations of change of data. The analysis is as follows:

Changes related to students: suspensions of schooling, resumption of schooling, dropout, dismissal, death, etc.

When students: suspend from school or resume schooling, they need to handle related formalities in the information center, entering related data into personnel data center system and sending the information to related departments through the system.

When students: dropout or are dismissed, related information will be sent to the smart card center, and the Student Union will come to handle related formalities(maybe not). The smart card center sends related data via the system, and set a deadline for students to handle formalities of leaving school. If the students do not come back within the deadline, it is necessary to report the loss of smart card manually and send information to related departments via the system.

Changes related to teachers: enrolling, transferring, retirement, death, etc.

When teachers: enroll or transfer into, they should handle related formalities in smart card center and input the change of data into the personnel center and send the information to related departments via the system.

When teachers: pass away, they should transfer data to the smart card center via personnel department and retirement department. The smart card center would then cancel the smart card and inform the personnel center the change, and transfer the data to the related department via the system.

Temporary workers: would enroll or leave.

When they enroll, they would get a smart card from the smart card center, and input the data into the personnel center. When they leave (when the card is invalid), the system would delete the smart card automatically, and send the changed information to the data center.

#### Data Archiving

As for the graduation, dropout, dismiss and death of students, the enrollment or death of teachers, and the quitting of the temporary personnel, the personnel center would archive all the data.

### **3 Building Demission Procedure System on the Basis of Campus Smart Card System**

As the basis of personnel data center, smart card is responsible for offering personnel data to all departments. Each year the personnel data of all departments would change because of the graduation of students and the retirement of teachers. How to simplify the process and provide unified and precise personnel change data is by now the issue to be researched and resolved.

We independently develop this demission procedure system on the basis of campus smart card center. Students could handle formalities with smart card in related departments instead of transferring and installing data in smart card and data center. Information would be presented on the internet, so school leaders and all departments could view it at any time. The system has saved a lot of time and energy for teachers and students, and is widely praised. The demission procedure system is an important step to digital campus, and the process of transforming the paper-based demission process into a digitized, cyberized and informationized one, and solidly fosters the building of the personnel data center for the whole campus.

## System Structure as Figure 1

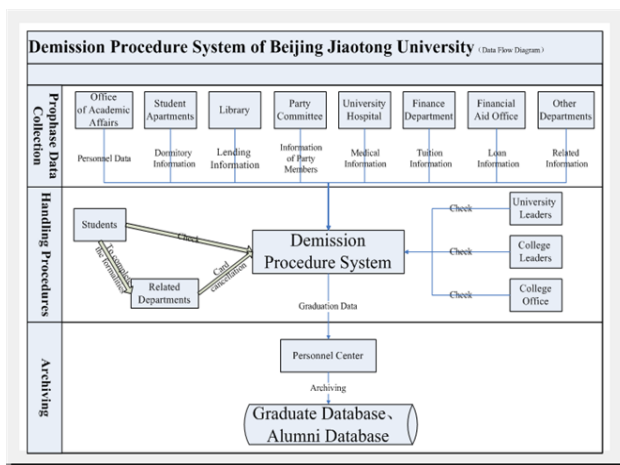


Fig. 1. Structure of Demission Procedure System

### Introduction to the System Structure

The system is divided into three parts: preliminary data preparation, demission procedures and data archiving.

#### One, Preliminary Data Preparation Phase

Preliminary data preparation phase: mainly responsible for inputting the information of personnel and departments. The departments related to leaving school involve Library (Students who have not returned books or have debts should handle related formalities), Student Department Centre (Students who rent room should handle related formalities), Party (Party member should handle related formalities), Finance Department (Students who are on debt should handle related formalities), Financial Assistance Office (Students who get a loan but have not signed repayment agreement), Teaching Subjects Office (Students who are on debt should handle related formalities), Energy Office (Some late-pay dormitory should pay utility fee), Information Centre (Pay internet fee and return switch board). These departments would input related information before students leaving school. There are some points worthwhile being noted: as for the data of Finance Department and Financial Assistance Office that are relative steady, there is no deadline; as for the data of Energy Office, Library, Information Centre which change rapidly, there should be immediate input before handling formalities. After data input, different departments have different approach: as for the fee of Energy Office and Information Centre, the university would pay; after Library has offered data graduates would not be able to borrow books any longer.

#### Second, Phase of Demission Procedures

Phase of demission procedures: Students log Information Management System, and search the departments that need handling formalities. There are three stages: to be handled, handled and no need. After students have completed related formalities in all departments, the departments would mark the smart card with 'handled'. Smart

card is the carrier of the process. The related data would be stored in smart card centre. As students cancel smart card, the mark “Alumni” would be printed on it and the card becomes an alumni card.

Third, Archiving Phase

Ailing phase: data would be stamped and filed through system and archived as alumni data.

How to handle formalities

Students should log Information Management System, search the departments that should handle related formalities and complete the process. Figure 2:

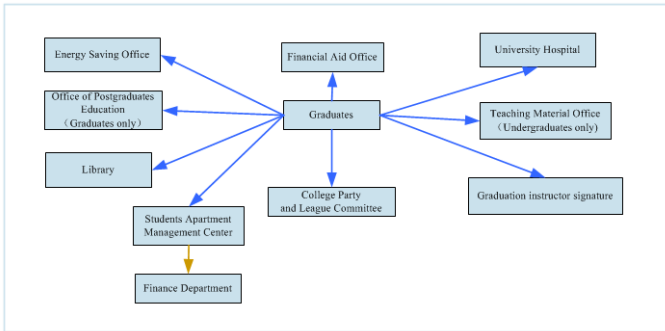


Fig. 2. Procedures of Completing the Formalities

## 4 Building Student Conducts Analyzing System on the Basis of Campus Smart Card System

Campus Smart Card system has taken place of different cards for effective identification and intellectual electric wallet, with the function of “one card for multipurpose, one card for general use”. Campus smart card system has run about two and half a year in our university now with 120 million turnover and 150 million data records. In order to use these data to provide decision support, we have developed the analytical system for student behavior based on campus smart card.

Main Functional Modules of the System

Data Transferring and Classified Split

In order not to affect the normal operation of the campus card system, we have designed and developed the data transmission service, which is scheduled to run in the background. From time to time, new data in the database will be recalled and sorted, dumping into the runtime later. In addition, basic data synchronization (customer information, unit information, etc.) can also be done here. Thus, data can be dumped to a query server, which can be published on the campus network for query and statistics.

Query Function

Query can be done for behaviors like recharge records, consumer records, dining records, computer records and access records under the choice of specific groups (such as a particular college, grade or students of a period).

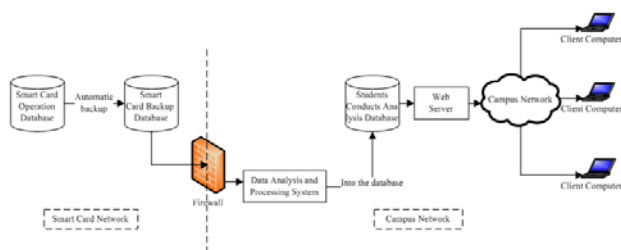
### Statistics Function

Statistic analysis and corresponding pie and bar graph can be drawn through recharge records, consumer records, dining records, computer records and access records under the choice of specific groups (such as a particular college, grade or students of a period).

### Notification Function

For each type of behavior analysis, the results will be recorded in the alert table, such as people who came to dormitory late for five times a month; who had meals for more than 60 times a month, yet paid less than 100 RMB; who was online for more than 150 hours a month, etc. These issues needed to be reminded will be recorded into the alert table. When university faculties who are responsible for management log in, those issues will pop-up reminding windows, and send to them in a way of text messages.

System Structure as Figure 3:



**Fig. 3.** Students Conducts Analysis Databaser mode

### Difficulties of the Campus Smart Card System

Currently, the flow data records of campus card system is up to 150 million. Query efficiency is the primary concern in such a large data for query, statistics, calculation and analysis. Thus, we adopted a variety of measures to optimize the query speed.

#### Data Sources Optimization

In order not to affect the normal operation of the campus card system, we do not collect data from the runtime library of the campus smart card, but from the backup server. The backup server backs data at four o'clock in morning every day according to the backup strategy. Meanwhile, it will restore data automatically through scripts. The data processing system will collect the previous day's data from backup database at eight o'clock in the morning, analyzing and summarizing the data. It takes 2 hours to collect 130,000 data everyday on average. The data of such behavioral analysis system lags behind one day than that of campus smart card system. However, the one-day lag is totally acceptable for the query statistical system.

#### Database Optimization

In order to improve the access efficiency, we create an index to the database, properly refer to some of the redundant data, and reduce the associated table to optimize the access efficiency. Meanwhile, we create some tables for statistics, which is operated regularly in the background by the stored procedure. It has directly access to the statistical results so as to achieve optimization.

### Algorithm Optimization

We have analyzed and summarized the raw data of campus smart card system, which records all the swiping card behavior. Thus, in the analysis of student behavior, firstly, we'll remove card records of external personnel; secondly, we have summarized dining data, access data, shower data, medical data from the original database when they were extracted. Take dining data as an example. Firstly, we summarize data in terms of meals, which means repeatedly clicking card in one meal corresponds to multiple records in the database, grouped into one meal; Secondly, statistics will be grouped by the amount and number of breakfast, lunch and supper. It will be faster to get the statistical graph. Thirdly, the reminded data will be recorded into reminder sheet(such as dining data that is over 70 times and less than 150 RMB ) rather than calculate each time so as to improving access efficiency.The result as Figure 4.



Fig. 4. Student Conducts Analyzing System Result

## 5 Conclusion

Campus smart card system is basic supporting part in campus information constructions. It is one of the projects which can best reflect the effectiveness of campus information construction. It also plays a great role in improving school management and service. According to practical needs, the way that integrates campus smart card system with other operation departments is encouraged. Furthermore, campus smart card system has been an inseparable part of digital campus constructions via extending its attributes to the whole university in order to provide bottom data support for digital campus constructions.

## References

1. Jing, Y.-W., Zhang, W.: Digital campus Facilitating University Construction Integration. China Education Network, 59–60 (April 2008)
2. Shen, X.-C., Chen, H.-C.: University information system specifications. Tsinghua Univ. (Sci. & Tech.) 43(4), 529–531 (2003)
3. Duan, Z.-M., Wang, R.-L., Sun, M.-Q.: Research and Implementation of the Digital Campus Resource Integration Based on Packaged Campus Cards. Computer Engineering & Science 30(1), 8–11 (2008)

# Research about Pattern Matching Algorithm

Minjie Wang and Lianxuan Zhu

College of Sciences, Henan Agricultural University Zhengzhou, China

**Abstract.** Several algorithms including BM, BMG, AC and AC-BM are discussed and the running time of the algorithms are measured on the snort in the paper. The results show that AC and AC-BM are faster than BM and BMG on the large number of patterns, but on the small number of patterns we obtained the opposite result.

**Keywords:** pattern, intrusion detection, string matching, algorithm.

## 1 Introduction

As one of the methods to detect and protect network security, intrusion detection system has been extensively applied in various network environments. Based on matching algorithm, intrusion detection system[1] aim at every intrusion activity to extract its characteristic value and write detection rules down according to the specifications, thus rule database has been generated; The captured data packet and rules in rule base were matched one by one, if matching could be made, intrusion activity was proved and alarming information would be output; otherwise, it would be proved to be normal and normal data packet would be discarded to release the internal storage. As the key of the intrusion detection system, matching algorithm directly influence the current performance of the system, therefore, to improve the performance of detection algorithm has been the hot issue of researching.

## 2 Definition of Pattern Matching

Definition of pattern matching of character string: on the character set  $\Sigma$ , a text character string  $T[1\dots N]$  with a length of  $N$  and pattern character string  $P[1\dots M]$  with a length of  $M$  are set.  $K$  is selected to mean the amount of pattern set, and  $P=\{p_1, p_2, \dots, p_k\}$  is selected to mean the pattern set. In case of  $1 \leq S \leq N$  and existing  $T[S+1\dots S+M]=P[1\dots M]$ , pattern  $P$  appears in the bit  $S$  of the text  $T$ , in another word, the pattern is matching the text. The problem of character string matching is to find whether  $P$  would appear in  $T$  or not and the appearing location.

E.g. Text character string  $T$ : He is a good boy.

Pattern character string  $P$ : good

Pattern character string  $P$  need to be searched in text character string  $T$ : "good". We found that pattern character string  $P$ : "good" appeared in the ninth place of text character string  $T$  by searching, namely, pattern character string  $P$  successfully matched the text character string  $T$ .





B. BMG [3] algorithm

BMG algorithm is the improvement based on BM, and this algorithm mainly regards uniqueness of latter character of the string as well as location difference between ending string and subsequent character of the corresponding text string in the pattern string. BMG algorithm uses the next character of corresponding character with the last location of text string and pattern string as inspiration to move right. The main idea of this algorithm is: when  $T[i+j] \neq P[i]$ , it means this round of matching is not made, then you need to see whether  $T[j+m]$  appears in pattern string uniquely or not.

- If  $T[j+m]$  does not appear in pattern string, it means  $T[j+m]$  does not match with any location of pattern string, and you could move the pattern string right for  $m+1$  bit;
- If bit  $T[j+m]$  just appears once in pattern string, and the right shift quantity corresponding to bit  $T[j+m-1]$  is bigger than bit  $T[j+m]$ , then pattern string could be moved right for  $m+1$  bit;
- If bit  $T[j+m]$  appears several times in pattern string, then the right shift quantity need to be conformed according to the amount of right shift bit corresponding to  $T[j+m-1]$  and  $T[j+m]$ .

Still for the text:

T: efbmherkadrhtenaf rderegrdad bher

Pattern P: dad bher

The matching process using BMG algorithm is shown as tab.2.

We will find that the comparison times using improved algorithm are 3 times fewer than primary algorithm, thus the pattern matching speed has been improved.

**Table 2.** Matching process using BMG

T	e	f	b	m	h	e	r	k	a	d	r	h	t	e	n	a	f	r	d	e	r	e	g	r	d	a	d	b	h	e	r
P	d	a	d	b	h	e	r																								
2										d	a	d	b	h	e	r															
3																															
4																															

C. AC[4] algorithm

Compared to the single pattern algorithm, AC algorithm belongs to the multi-pattern matching algorithm. The advantage of multi-pattern matching algorithm is that multiple patterns could be matched just by one traverse. At present, the classic multi-pattern matching algorithms include AC and AC-BM algorithm.

Idea of AC algorithm: in the stage of preprocessing, three functions including swerve function, failure function and output function will be built by finite automata algorithm, thus a tree finite automaton could be created. The processing procedure matching by this pattern turns to be processing procedure of state transition beginning with “0” state. In the searching stage, these three functions are crosswise used to scan text and position all of the appearing locations of key words in text.

Here, example is given to illustrate the processing procedure using AC algorithm. For example, finite automaton is used to search pattern string of {he,she,his,hers} in the target string “ushers”:

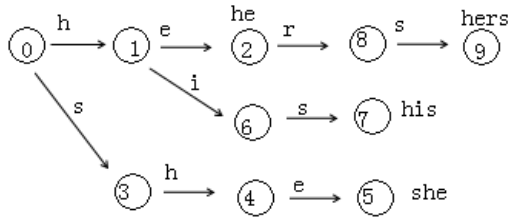


Fig. 1. Swerve function

A automata of tree type as shown in Fig.1 is generated by finite automata algorithm, state 0 is the initial state, the set of this state is {0,1,2,3,4,5,6,7,8,9}, and the input set is character string of Chinese or English, output set is pattern string set {he,she,his,hers}.

Swerve function  $G(s, t)$  is a function that the character matching of one text character string is performing from one state to another. If it reaches a ending state, it means pattern designated by output function has been matched successfully. And it is explained by the arrows and character above in Tab.3, and when it is defined at the initial state  $s$  of arrows, if the “next” input is character  $t$  on the arrow, then the state of automata will turn to  $G(1, e)$ , e.g.  $G(1, e)=2$ .

Table 3. Output function

State	Output
2	he
5	she
7	his
9	hers

Failure function  $F(s)$ : Failure function means that when it is at the stage  $s$ , if the next input  $t$  do not exist on the arrows of any swerve function, in another word, one input character causes the mismatching, then the state of automata will turn to  $F(s)$ . Failure function is used to select a state; in addition, matching will be started again from this state.

Output function  $P(s)$ :  $P(s)$  means the applicable output provided by automata at stage  $s$ , it means the finding target string.

Therefore, the process of searching pattern character string in text character string has been transformed to process of searching in pattern tree. The searching of one text character string  $T$  should begin from the root node of pattern tree, and go down along the path which takes characters in  $T$  as label; if the automata can reach the ending state  $v$ , then it means that there exists pattern  $L(v)$  in  $T$ ; otherwise, no pattern exists.

#### D. AC-BM algorithm

AC-BM algorithm is a new algorithm raised by combining and improving AC algorithm and BM algorithm by Jang-long, and this algorithm combines multiple patterns in need of searching as structure of tree type, and selects the same prefix as root node of tree, so this pattern tree is named multi-pattern rule tree, which is also known as key word tree. The pattern tree moves from the right of text to left, and once

it ascertains a proper location, character comparison will begin from left to right. This algorithm depends on the derived function of heuristic function of BM algorithm, and what is different from BM algorithm is that just one pattern could be moved by BM algorithm, in contrast, by AC-BM algorithm, a pattern tree composed by all patterns could be moved, and meanwhile, bad character and good prefix are used to move. The movement of bad character is similar with BM algorithm, but some modifications have been made. The detailed methods are as follows: if mismatching happened, pattern tree should be moved to make the other patterns in the tree match with the character which is matching with the comparing character in current text so as to perform matching. If in the current depth, there is no text character appearing in any pattern, then the shift quantity of pattern tree is just the length of the shortest pattern in the tree. The methods of good prefix are: pattern tree is moved to the next location of completely prefix of substring of another pattern which has been found, or moved to the next location where the suffix of another pattern in the tree could correctly match with some prefix of the text. During the movement of the pattern tree, make sure that the shift quantity of pattern tree should be not larger than the length of the shortest pattern in the tree.

## 4 Experiment

This paper experimentizes in the experimental platform of CPU type: Doalcore AMD Athlon64\*2, 2300MHz(11.5\*200)4400+ ; Memory: 256M; Hard Disk: 8.0G; System: fedora10; Snort version: snort 2.8.6 to compare the matching time of pattern in BM algorithm, AC algorithm and AC-BM algorithm respectively. To reduce the influence caused by network sniffer time to the experiment results, 10 M data of local log collected by the Snort system of this machine is used for experiment data set, and the amount of pattern in the test are 10, 20, 50, 100, 200, 400, 800 respectively, which are generated randomly by programs. The experiment results are as shown in Fig 2.

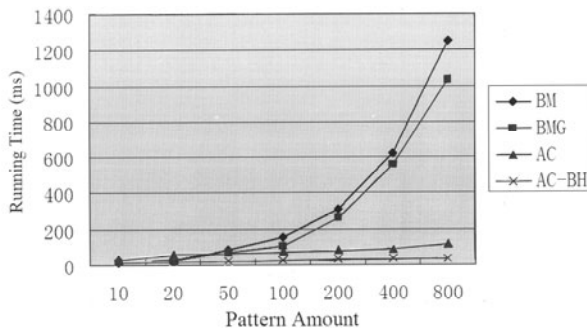


Fig. 2. Running time comparison of matching algorithm with the increasement of pattern amount

## 5 Results Analysis

(1) The matching speed of BM and BMG algorithm is faster in the situation of fewer pattern sets, in another word; single pattern algorithm is suitable for the matching of small pattern set.

(2) With the increasement of pattern amount, the increasement of running time of algorithm matching is little. Referring to the reasons, it mainly because the increasement of pattern amount enlarges the scale of matching automata structured by pattern set, and then the automata needs to spend longer time on shifting state.

## 6 Conclusions

With the improvement of network speed and enlargement of network data flow, matching algorithm has been the key of intrusion detection system, and the speed of character string matching has become the bottleneck to apply intrusion detection system. Therefore, further research of pattern string matching algorithm is significant. This paper discussed four kinds of algorithms including BM, BMG, AC, AC-BM to match character string pattern, and measured the running time of this four algorithms in Snort. The results showed that when the pattern amount is larger, algorithm AC and AC-BM are better than BM, BMG; when pattern amount is fewer, the result is just the opposite.

**Acknowledgments.** This paper is funded by natural science fund project of provincial department of education in Henan: Intrusion detection research based on protocol analysis (2009B510010).

## References

1. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of CVPR, Kauai Marriott, Hawaii, pp. 511–518 (2001)
2. Freund, Y., Sharpie, R.E.: A decision-theoretic generalization of online learning and an application to Boosting. *Journal of Computer and System Sciences* 55(1), 119–139 (1997)
3. Friedman, J.H., Trevor, R.T.: Additive logistic regression: A statistical view of boosting. *The Annals of Statistics* 38(2), 337–374 (2000)
4. Aho, A.V., Corasick, M.J.: Efficient String Matching: An Aid to Bibliographic Search. *Communications of the ACM* 18(6), 333–343 (1975)
5. Commentz-Walter, B.: A string matching algorithm fast on the average. In: Proc. 6th International Colloquium on Automata, Languages, and Programming, pp. 118–132 (1979)

# Study of E-mail Filtering Based on Mutual Information Text Feature Selection Method

Shangfu Gong, Xingyu Gong, and Yuan Wang

School of Computer, Xi'an University of Science and Technology,  
Xi'an 710054, Shanxi, China  
gongsf@xust.edu.cn

**Abstract.** Aiming at the problem with filtering E-mail, based on analyzing defects of the traditional mutual information, an approach based on quadratic  $TF * IDF$  mutual information feature selection is presented in the paper; then the importance of characteristic words appearing just in only one class is again measured to solve the problem that feature selection is not effectively done because of equal mutual information value. Finally, Bayesian classifier is used for experiment and experimental result shows that compared with the original method, the presented approach possesses higher correct rate and more efficiency of classification in text classification.

**Keywords:** E-mail Filtering, feature selection, text classification, mutual information.

## 1 Introduction

E-mail in internet has become the important way to transmit information during work, live and contact. However, the emergence of a large number of spam and interference causes great harm to work efficiency and information security. How to filter Spam effectively has become one of the important issues. Usually, the vocabulary made up the message is very large. According to the results of Paul JJ Payack, Global Language Monitor president and chief word analyst, the number of English words has reached 1,005,939. A large vocabulary is a challenge for filtering spam based on message content and the predecessors in this matter has done a lot of research and experiment [2,3].

Feature selection algorithm is one of the effective ways to solve such problems. It can reduce the text of the feature vector dimension, remove redundant features, and retain differentiated features. There are several commonly feature selection algorithm [4]: document frequency, information gain,  $\chi^2$  statistics, mutual information, low-loss dimension reduction and frequency difference method. In filtering spam, enough training samples and high coverage of words are helpful for effective classification of spam. If feature selection is poor, dimension of feature library is large and time taken for searching feature words. In this paper, the mutual information is applied for feature selection of text of English mail.

## 2 Mutual Information Algorithm

### A. Traditional Mutual Information Algorithm

In information theory, mutual information is used to measure the degree of correlation between two piece of message[5]. In feature selection, the mutual information is used to characterize the associated degree between feature  $t$  and class  $C$ . The greater the mutual information is, the greater the co-occurrence degree of feature  $t$  and text class  $C$  is. The mutual information between feature  $t$  and text class  $C$  is defined as formula 1.

$$I(t, C) = \log \frac{P(t, C)}{P(t) * P(C)} \quad (1)$$

In formula1,  $P(t, C)$  means the co-occurrence probability between feature  $t$  and class  $C$ ,  $P(t)$  means the frequency to feature  $t$  appearing in the training set, and  $P(C)$  means the probability to class  $C$  in the training set.

There are two statistical methods of traditional mutual information to feature  $t$  and the entire message set: the maximum mutual information and the average mutual information method. They are defined as formula 2 and 3.

$$MI_{\max} = \max_{i=1}^m I(t, C_i) \quad (i=1,2,\dots,m) \quad (2)$$

$$MI_{\text{avg}} = \sum_{i=1}^m P(C_i) * I(t, C_i) \quad (i=1,2,\dots,m) \quad (3)$$

### B. Defects of the Traditional Mutual Information

In all of the derivation formula, assuming that the number of spam and the normal e-mail is equal in training set.

$$\begin{aligned} MI_{\max}(t) &= \max_{i=1}^2 I(t, C_i) = \max_{i=1}^2 \log \frac{P(t, C)}{P(t) * P(C)} \\ &= \max_{i=1}^2 \log \frac{P(t | C)}{P(t)} = \max_{i=1}^2 \log \frac{D_{ti} * D_{All}}{D_i * D_t} \\ &= \log 2 * \max_{i=1}^2 \log \frac{D_{ti}}{D_t} \end{aligned} \quad (4)$$

In formula 4,  $I(t, C_i)$  means mutual information between feature  $t$  and class  $C_i$ ,  $D_{all}$  means the number of all the mail set,  $D_{ti}$  means the message number of feature  $t$  contained in the class  $C_i$ ,  $D_t$  means the mail number of feature  $t$  contained in the mail set,  $D_i$  means the number of mail in the class  $C_i$  and  $D_{all} = 2 * D_i$  is defined.

In the maximum mutual information, because of  $D_{ti} / D_i < 1$ ,  $\log(D_{ti} / D_i) < 0$  is true. According to formula 4, if feature  $t$  only appear in one class, the mutual information reach maximum value:  $MI_{\max(t)} = \log 2$ .

$$\begin{aligned}
MI_{avg} &= P(C_1) * I(t, C_1) + P(C_2) * I(t, C_2) \\
&= \frac{1}{2} \left( \log \frac{P(t|C_1)}{P(t)} + \log \frac{P(t|C_2)}{P(t)} \right) \\
&= \frac{1}{2} \left( \log \frac{D_{t1} * D_{All}}{D_1 * D_t} + \log \frac{D_{t2} * D_{All}}{D_2 * D_t} \right) \\
&= \left( \frac{1}{2} * \log 2 \right) + \left( \log \frac{D_{t1}}{D_t} + \log \frac{D_{t2}}{D_t} \right)
\end{aligned} \tag{5}$$

$D_{t1}$  means the number of mails containing feature  $t$  in spam set and  $D_{t2}$  means the number of mails containing feature  $t$  in normal mail set.

In average mutual information, according to the logarithmic functions of the formula 5, both  $D_{t1}$  and  $D_{t2}$  are less than  $D_t$ , hence  $\log(D_{ti}/D_t) < 0$  is true. When the feature  $t$  just appears in only one class, namely  $\log(D_{ti}/D_t) = 0$ , the mutual information reach the maximum value:  $MI_{avg(t)} = (1/2) * \log 2$ .

Based on the above analysis, using the two mutual information methods, there are some obviously defects. According to the threshold to reduce the number of feature words, the larger amount of mutual information must be selected. But for some feature words (that is, independent of a class and related to another class), when using these two methods, mutual information value is a fixed and maximum value. Hence, in the case of the small training set, due to too large number of English vocabulary and training set is limited, a large number of words in only one class will appear in training set. Traditional mutual information method can not judge the important degree between them, and then can not pick out the feature words that contribute to the classification. Thus, a more efficient way must be found out to measure their mutual information value.

### C. Mutual Information Feature Selection Based on Quadratic TF \* IDF.

*TF \* IDF* (Term Frequency \* Inverse Document Frequency) is a statistical method for evaluating the importance of a word for a document set or a document for a corpus. The higher the frequency of a word in the document, the better the distinction ability of the content attributes (TF); the wider the range of a word in the document, the worse the distinction ability of the content attributes (IDF) [6,7]. Formula 6 shows that the use of TF \* IDF to calculate the weight of feature  $t$  to document  $j$ .

$$W(t_j) = TF * IDF = ((\log TF) + 1) * \left| \log \frac{D_{All}}{D_{t1} + D_{t2}} \right| \tag{6}$$

$$W(t) = \sqrt{\sum_{j=1}^n (W(t_j))^2} \quad (j=1,2,\dots,n) \tag{7}$$

In the formula, TF represents the frequency of feature  $t$  appearing in all the mail sets. IDF represents reciprocal of document frequency of feature  $t$  appearing in the all mail sets.

On one hand, quadratic TF \* IDF conquers the defects that traditional mutual information did not consider the frequency, on the other hand, it enhances the ability of traditional method to measure feature words appearing in only one class. As the feature



set is the best character to distinguish classes in the all training sets, the weight of feature  $t$  for all documents must be calculated. However, the traditional method just calculates  $TF * IDF$  weight of a single document, To solve this problem, the formula 7 is proposed to calculate the weight of feature  $t$  for all documents.  $MI = MI_{\max} + W(t)$  is defined as the mutual information value of the feature appearing only in one class.

#### D. Description of Algorithm

Algorithm: the algorithm based on quadratic  $TF * IDF$  mutual information feature selection

Input: training set  $T = \{t_1, t_2, \dots, t_{all}\}$ , class set  $C = \{c_1, c_2\}$ ,  $c_1$  represents spams,  $c_2$  represents normal mails,  $n$  represents the minimum frequency,  $w$  represents the minimum document frequency,  $\min$  represents the minimum value of mutual information,  $\max$  represents maximum mutual information value .

Output: feature set  $L = \{l_1, l_2, \dots, l_k\}$

BEGIN

Make statistics of all the mails of training set  $T$ , including term frequency and document frequency stored to the array  $word[i]$ .

WHILE  $i < word.length$

DO

- 1) If  $DF_i > n \ \& \ TF_i > w$ , then  $character = word[i]$ ;
- 2) use formula 2 or formula 3 to calculate the mutual information value  $MI$  of the character
- 3) If  $MI = \max$ , then  $MI = \max + W(character)$  ;  
//  $\max$  is maximum mutual information  $MI$ ,  $MI_{\max} = \log 2$  or  $MI_{avg} = (1/2) * \log 2$   
use formula 6 or formula 7 to calculate weight of the character
- 4) If  $MI > \min$ , then  $L = character$  ;

END DO

END

### 3 Experimental Design and Performance Analysis

#### A. Sample Set

In experiment, Ling-Spam Corpus is applied . It consists of spam and the non-spam of Linguist list and is divided into 10 parts, including the 482 spam sets and 2412 normal

sets, each one containing about 289 mails. It can be downloaded <http://labs-repos.iit.demokritos.gr/skel/iconfig/downloads/>. In the experiment, we take the top 5 sets of the lemm\_stop as the training sets, take the bare e-mail last 5 sets as the test sets.

## B. Standard of Evaluation

Six ordinary evaluation standards are used to evaluate the results [7]: Recall, Precision, Accuracy, Fallout, Error, and Miss rate.

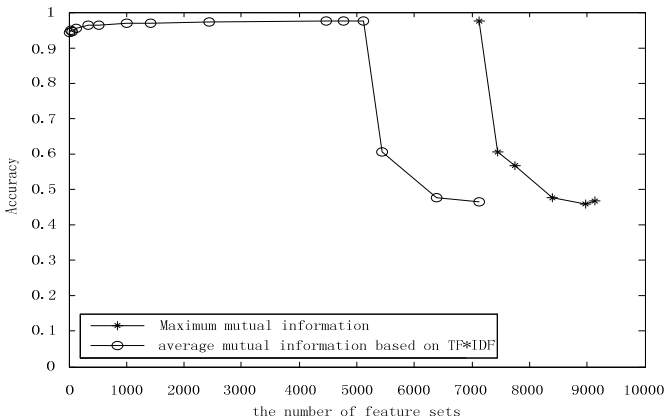
## C. Classifier

Bayesian classifier [8] is a statistical classification method, which can predict the likelihood of class membership, such as the rate of a given character to one class. For large database, the Bayesian classifier method also shows high precision and high efficiency, and is most sophisticated and common in content filtering technology. Therefore, Bayesian classifier is chosen in this study.

## D. Experiment

Before feature selection, all the mails in training set to remove the outdated words and word stems. Training sets include 400 mails, which contain 200 spam and 200 normal mails. Then, the words statistics reaches a total of 9139 in the training set. Test set consists of 1445 mails, including 241spam and 1204 normal mails.

Traditional mutual information and quadratic TF \* IDF mutual information are used,; then, the threshold is set to make feature selection for training set ; finally the test result is shown in Fig. 1 and Fig. 2.



**Fig. 1.** Comparing accuracy of traditional method with that of secondary TF\*IDF mutual information

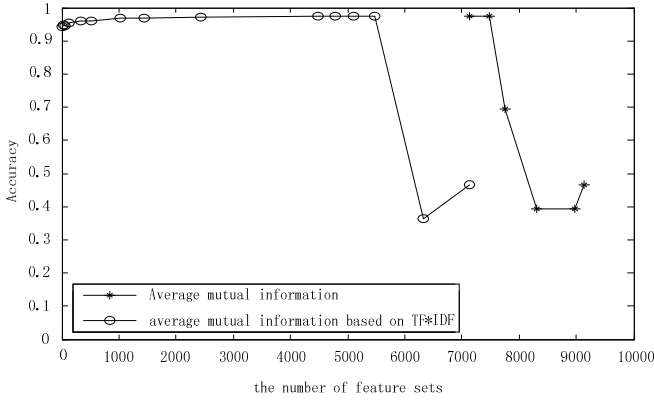


Fig. 2. Comparing accuracy of traditional method with that of secondary TF\*IDF mutual information

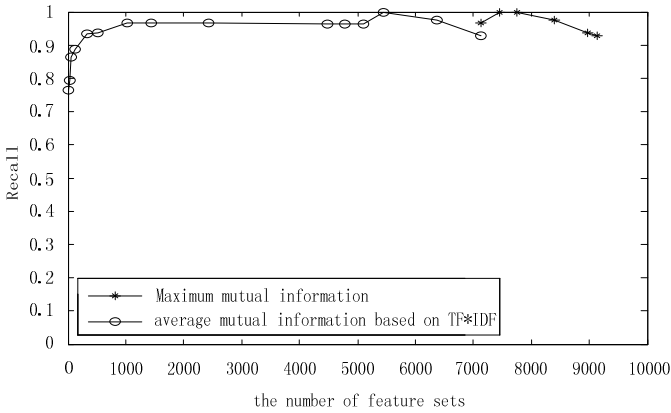


Fig. 3. Comparing recall of traditional method with that of secondary TF\*IDF mutual information

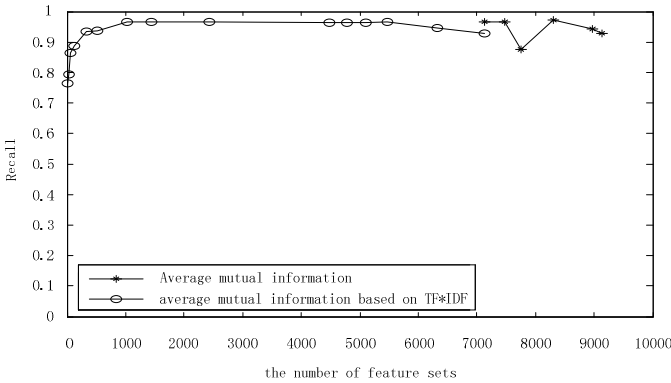


Fig. 4. Comparing recall of traditional method with that of secondary TF\*IDF mutual information

According to the experimental data of traditional mutual information, the dimension of feature set should be set as 7122 because large feature words appears in one class of mail set and single mutual information can't distinguish further their degree of importance. Accuracy maximum reaching 97.4% and 96.7% of recall rate, dimension reduction is not effective. In quadratic TF \* IDF mutual information method, equal mutual information is calculated again. Degree of importance of feature words can be distinguished according to their weight in whole document. Classification results are best when dimension of feature set is within 1000-5112, accuracy rate is between 96.9% to 97.6% and recall rate ranges from 96.3% to 96.7%. It not only keeps the original recall rate and raises classification precision, but also reduces effectively dimension from 7122 to 1000. In addition, on the other four evaluation standards, the quadratic TF \* IDF mutual information shows better results.

On the aspect of time efficiency, the 8th part of bare version including 239 mails is used as test set. After extracting the feature set by applying traditional method, Classification takes more than 195 seconds, but improved method only takes 53 seconds to 150 seconds . This is because lower-dimensional feature set can raise the query efficiency.

## 4 Conclusion

After studying traditional mutual information methods , It is found that they can not take good measures of the feature words appearing only in one class. Thus, by applying quadratic TF \* IDF to make the feature set obtain a better ability of classification. Experimental result shows that while after using quadratic TF \* IDF mutual information, classifier not only keeps the original recall, but also it's precision of classification rises to some degree.

## References

1. Payack, P.J.J.: Number of Words in the English Language: 1,005,939 (EB/OL) (July 4, 2010), <http://www.languagemonitor.com/>
2. Dasgupta, A., Drineas, P., Harb, B., Josifovski, V., Mahoney, M.W.: Feature Selection Methods for Text Classification. In: Proceedings of the 13th ACM SIGKDD International Conference, pp. 230–239 (2007)
3. Zheng, Z., Wu, X., Srihari, R.: Feature Selection for Text Categorization on Imbalanced Data. ACM SIGKDD Explorations Newsletter 6(1), 80–89 (2004)
4. Zhang, H., Wang, L.: Automatic text categorization feature selection methods research. Computer Engineering and Design 27(20), 3838–3841 (2006)
5. Lu, Y.: Reserch on Content-based Spam Filtering Technology. Southwest Jiaotong University (2009)
6. Shi, C., Xu, C., Yang, X.: Study of TFIDF algorithm. Journal of Computer Applications 29, 167–170 (2009)
7. Huang, Z.: Design and Implement of Chinese-based Bayesian Spam Filtering System. University of Electronic Science and Technology of China (2007)
8. Han, J., Kamber, M.: Data Mining Concepts and Techniques. China Machine Press, Beijing (2008)

# Research on the Moving Vehicle Detection Algorithm Based on the Motion Vector

Jian Gong, Fuqiang Liu, Chunlin Song, Jianzhu Cui, and Zhipeng Li

School of Electronics and Information Engineering, Tongji University,  
Shanghai 201804, China  
gjguitar2008@163.com, fuqiangliu@163.com,  
songchunlin@mail.tongji.edu.cn, cuijianlou@163.com,  
newspring88@126.com

**Abstract.** In this paper, moving object detection under the dynamic background in aerial videos has been studied, and a moving vehicle detection algorithm based on the motion vector is proposed, using the KLT algorithm for matching feature points, according to K-medoids clustering algorithm on the feature points, finally the location of the target vehicle has been found. Experiments show that the method in this article can accurately detect the moving vehicles in the aerial videos, and then extract the corresponding traffic information.

**Keywords:** moving target detection, KLT feature matching algorithm, K-medoids clustering algorithm.

## 1 Introduction

Moving object localization in aerial video is one of the hotspots in the research of computer vision field, no matter in the military or business goal orientation, the application value of the technology has been recognized. In recent years, along with the proposal of background motion compensation technology, people gradually start deep study in target localization technology under the moving background. Since aerial images are the image sequence which photographed in the flying platform, the movement of flying platform and lens would cause the irregular background movement in the image. By the constraint of shooting conditions, as well as the influence of illumination and noise, there are more obstacles to the localization of moving objects in aerial image.

In 2002, the image processing and pattern recognition institute in Jiao Tong University, put forward a kind of parameters transforming method which made use of huber function and even sparse sampling estimation [1], to detect and locate the vehicles in aerial images and tried to compensate for the interference of background motion and noise on target. But the algorithm was very sensitive to light, and after the moving background compensation there would still be part of background unable to eliminate, which was difficult to object localization. In 2007, by analyzing the relative movement relationship between image, imaging flat and ground scenery, people established an aerial image motion blur model, analyzed the relationship between the

fuzzy zone width, the camera movement and the exposure time, and designed the fuzzy zone width measurement algorithm based on fuzzy zone edge strengthen and autocorrelation analysis, based on which they proposed the speed measurement method in image base on the motion blurring effect [2]. This method didn't increase additional equipment and ensure data, and it was easier to realize. The example showed that the algorithm for speed without a drastic change in the prospective uniform motion had good effect, providing a solution for the flight vehicle to measure the real-time flight velocity. In 2008, Zhe Jiang University aimed at the aerial image with complicated scene and proposed a kind of multi-resolution image matching algorithm based on the chart division theory and the Hausdrff Distance [3], which was helpful to match complicated scene image in a large deformation and strong noise interference. But during calculating the curvature radius, the error would be magnified in lots of complicated calculation. Besides, in this algorithm the border of image segmentation area was not consistent, leading to the inconsistent regional boundary extraction. According to the characteristics of aerial video this paper puts forward a moving vehicle detection algorithm based on the motion vector, it has a good effect on extracting vehicle information in aerial video, providing a more accurate basis for analyzing the traffic information in videos.

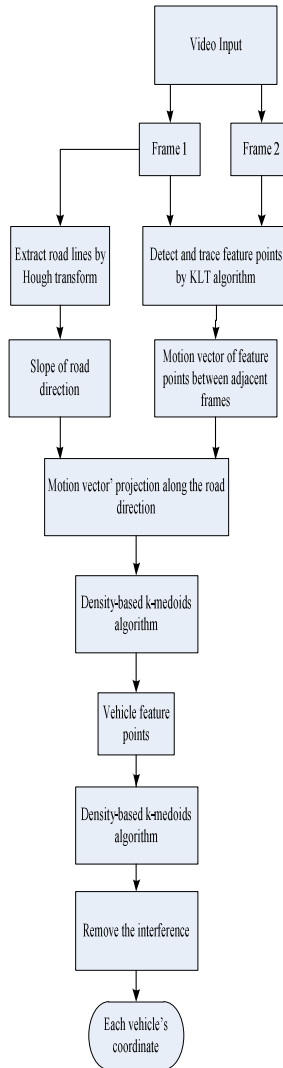
## **2 Moving Vehicle Detection Algorithm Based on the Motion Vector**

### **Algorithm Overview**

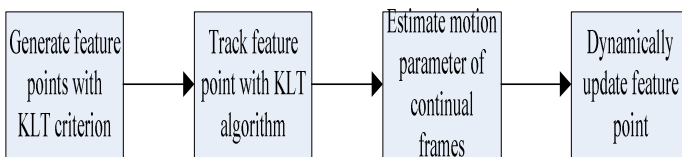
The common moving vehicle detection algorithm is based on edge detection and plaques analysis, which is proposed to be coarse, by large probability of mistake and omission in detection. This paper raises a kind of moving vehicle detection algorithm based on the motion vector, uses KLT algorithm to detect and track feature points, and gets the motion vector between frames, which corrects the feature points location, and then uses the K-medoids algorithm based on density to cluster, finally locates the position of vehicles. This algorithm can improve the detection precision and identify moving object in the image more accurately. Algorithm process shows as Figure1.

### **KLT Feature Point Matching Algorithm**

Under the condition of similar gray in feature areas between adjacent frame of image, KLT algorithm selects lots of feature points in gray-scale image sequence and takes a two-dimensional tracking, as a result, the location of feature points has been found, by which two-dimensional feature motion parameters can be calculated. The process of motion parameters estimation between frames using KLT algorithm is showed in Figure 2.



**Fig. 1.** Motion vector based moving vehicle detection algorithm process



**Fig. 2.** Process of motion parameters estimation between frames using KLT algorithm

KLT operator firstly calculates the feature value,  $\lambda_1, \lambda_2$  of the  $\delta$  matrix in each point, if  $\min(\lambda_1, \lambda_2) > H(\text{threshold})$ , then this point is an effective feature point. Delta matrix is defined as

$$\delta = \begin{bmatrix} I_x I_x & I_x I_y \\ I_y I_x & I_y I_y \end{bmatrix} \tag{1}$$

As the camera moves, the image strength changes in complex way. If the camera captures the image fast enough, then for the adjacent frames, as the similarity between kinds of factors that affect the gray level changes, the gray level changes similarly in the local area. So it can be considered that, there are displacement along the X direction and Y direction in the local area between two adjacent frames, which is called two-dimensional model of characteristic translational motion. This means that at time t the feature point  $X=(x, y)$  moves to the point  $X' = (x - d_x, y - d_y)$  at time t+1, in which  $d=(dx, dy)$  is the translational motion parameter vector of two-dimensional feature. The feature point's gray values are approximately equal before and after the movement, which is

$$J(X) = I(X - d) + n(X) \tag{2}$$

Obviously we must select the appropriate motion parameter vector d so that in the feature window W around the feature point X the following double integral gets the smallest residual,

$$\varepsilon = \int_W (I(X - d) - J(X))^2 \omega dX \tag{3}$$

Where  $\omega$  is a weighted equation to the pixels in feature region. If the movement between two adjacent frames is small, the expression  $I(X-d)$  can do first-order Taylor expansion at point X.

$$I(X - d) = I(X) - g \cdot d \tag{4}$$

G is the gradient vector, so we can rewrite equation (3) into the form below,

$$\varepsilon = \int_W (I(X - d) - J(X))^2 \omega dX = \int_W (h - g \cdot d)^2 \omega dX \tag{5}$$

Where  $h = I(X) - J(X)$ . It can be seen that residual is the quadratic equation of translation vector d, this optimization problem can get closed-form solution. In order to minimize the residuals, we get the first derivative of d on both sides of equation (5), then we get

$$\int_W (h - g \cdot d) g \omega dX = 0 \tag{6}$$



Since  $(g \cdot d)g = (gg^T)d$ , and  $d$  is assumed to be constant in the feature window area, so we get

$$d \times \int_{\omega} (gg^T) \omega dX = \int_{\omega} hg \omega dX \quad (7)$$

the above equation is the basic step for calculation on feature tracking, of all pixels in the feature window it can calculate the gradient along the X and Y direction, so we can get a real-symmetric cross gradient matrix  $G$ , meanwhile, of all the pixels in the feature window we can calculate the gray difference between two frames and get vector  $e$ . so we can calculate the value of feature parameter  $d$ . After that, move the feature window and repeat the process above until  $d$  is less than a certain threshold, which means the feature windows of two adjacent frames have been matched successfully, and the final translational motion parameters will be get by adding the values of  $d$  each round repeated above together.

### Road Line Extraction Method

Road line extraction can project the motion vector obtained by KLT algorithm along the direction of the road, reducing the image error caused by plane's jitter. In this paper, Hough transform is used to detect the road line. Hough transform can detect all the lines in the image, then we can get the slope  $K$  and intercept  $B$  of all lines and calculate the range of slope having the maximum distribution and then get the average value  $K_{avg}$ , by which we can obtain the direction of the road line and filter out the other unrelated lines. The maximum intercept  $B_{max}$  and minimum intercept  $B_{min}$  can be derived from the intercepts above and then the two borders of the road can be drawn, which can be used to roughly estimated roughly where the target area is. Figure 3 shows the diagram of the road line detection effect.

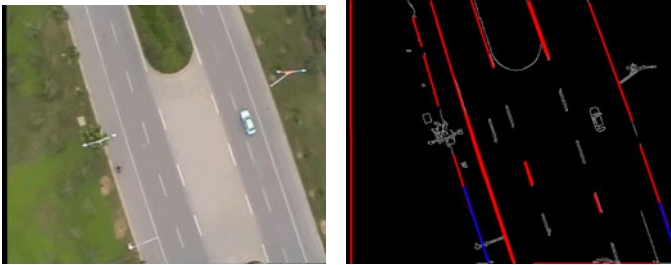


Fig. 3. Road line extraction(the blue ones are the road line)

### Density-Based K-Medoids Clustering Algorithm

K-means method is a common clustering algorithm, which requires inputting  $k$ , then  $n$  data objects are divided into  $k$  clusters so that it makes the cluster obtained meet: objects in the same cluster have high similarity; objects in different clusters have low similarity. Cluster similarity is calculated by a mean value called "central object" (gravity center) obtained from all clusters. However, this algorithm will not produce much difference between the size of clusters, and it is very sensitive to the dirty data,

besides, the number of vehicles needs to be inputted in advance, which is difficult to know beforehand in object detection. Therefore, this article uses a better algorithm: K-medoids method. Here we select an object called medoid to replace the central role above, so that a medoid identifies a cluster. The steps are showed as follows,

- 1) Freely select  $k$  objects as medoids( $O_1, O_2, \dots, O_i \dots O_k$ );
- 2) Assign the remaining objects to each category (according to the principle of the closest medoid);
- 3) For each class ( $O_i$ ), select an  $O_r$  by sequence, calculate the consumption- $E(O_r)$  of replacing  $O_i$  with  $O_r$ . Select the  $O_r$  with minimum  $E$  to replace  $O_i$ , then  $k$  medoids change;
- 4) Loop from step 2 until  $k$  medoids are fixed.

This algorithm is not sensitive to dirty data and abnormal data, but there is more amount of calculation than K-means algorithm, which means that it is generally suitable for small amount of data. So it is very suitable for the scene with small traffic just like the sparse traffic network involved in this paper.

Density-based K-medoids algorithm does not input the number of clusters, but merges the adjacent clusters according to the distance between data. This is a major improvement relative to K-means algorithm and also improvement in the program. After clustering by K-medoids algorithm, the location of the target vehicle feature points have been extracted and marked in the original image.

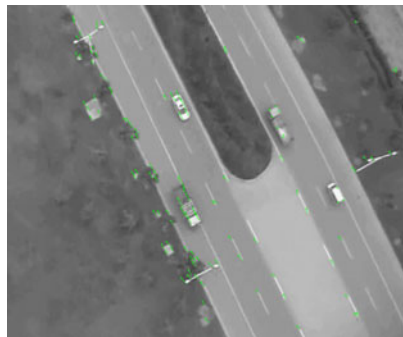
### 3 Experiment Results and Analysis

Algorithm in this article is written in C++ language, and runs in VC2005 compiler environment. Test PC uses Intel Pentium Dual 1.6GHz CPU and 2G memory. The experiments have realized the algorithm described in the paper and compared with common algorithms. Figure 4 is a schematic figure of detecting the feature points using KLT algorithm, Figure 5 shows the motion vector getting from comparison of two adjacent frames after getting feature points by KLT algorithm, Figure 6 shows the comparison of the effect on the algorithm based on edge detection and plaque analysis and the one proposed in this article, Figure 7 compares the effect between K-means clustering algorithm and K-medoids clustering algorithm. After several tests, the method proposed in this paper can reach an accuracy of 80 percent in detecting moving vehicles.

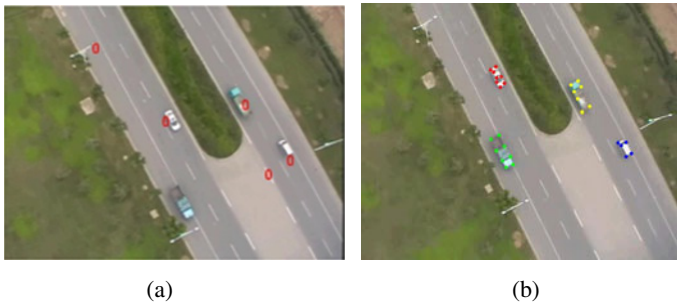
We can see that this method can detect the target vehicle more accurately, the extraction of traffic information is more accurate, and it has good practicality. During the test, we found that there are still some errors in the detection of road lines, and the detection of vehicle movement in the video is still missed sometimes when the video is frame-skipping, which are the focus of the future research.



**Fig. 4.** Feature points by KLT algorithm



**Fig. 5.** Motion vector getting by comparison between two adjacent frames



**Fig. 6.** Comparison between algorithm based on edge detection and plaques analysis(a) and the one based on feature points and clustering(b)



**Fig. 7.** Comparison between clustering algorithm based on K-means(a) and the one based on K-medoids(b)

## 4 Conclusion

A moving vehicle detection algorithm based on motion vector is proposed in this paper, it realizes the detection of moving objects in dynamic background using KLT matching algorithm and K-medoids clustering algorithm. The method is able to detect moving vehicles in aerial video better and detection accuracy up to 80%, so it provides accurate vehicle information for the analysis in sparse network traffic.

**Acknowledgments.** This research is supported by National Nature Science Foundation of China (60904068) and 863 Project from Science and Technology Department (2009AA11Z220), all support is gratefully acknowledged.

## References

1. Xu, D.-W., Gu, L., Liu, C.-Q.: Moving vehicles detection and location in aerial image sequence. *Computer Engineering* 28(9), 27–29 (2002)
2. Xu, C., Liu, Y.-C., Qiang, W.-Y., Liu, H.-Z.: Image velocity measurement based on motion blurring effect. *Infrared and Laser Engineering* 37(4), 625–650 (2008)
3. Chen, S.-Y., Qian, H., Wu, Z., Pan, L.-L., Zhu, M.-L.: Multiresolution image matching method based on graph-cut and Hausdrff Distance. *Journal of Image and Graphics* 13(6), 1185–1190 (2008)
4. Nieto, M., Arrospe, J., Salgado, L., Jaureguizar, F.: On-board video based system for robust road modeling. *Content-Based Multimedia Indexing*, 109–116 (2008)
5. Lv, C.-L., Wang, Y.-J.: An object motion prediction method combined with background motion estimation. *Opto-Electronic Engineering* 36(5), 11–16 (2009)
6. Sun, Z.-H., Kong, W.-Z., Zhu, S.-A.: Moving object location in viedo sequence based on subtractive clustering algorithm. *Opto-Electronic Engineering* 35(7), 12–16 (2008)
7. Hu, S.-J., Ge, X.-W., Chen, Z.-H.: Based on corner feature KLT track panoramic mosaic algorithm. *Journal of System Simulation* 19(8), 1742–1744 (2007)

8. Ma, S., Wang, T.-J., Tang, S.-W., Yang, D.-Q., Gao, J.: A fast clustering algorithm based on reference and density. *Journal of Software* 14(6), 1089–1095 (2003)
9. Tao, X.-M., Xu, J., Yang, L.-B., Liu, Y.: Improved cluster algorithm based on K-means and particle swarm optimization. *Journal of Electronics & Information Technology* 32(1), 92–97 (2010)
10. Chai, D.-F., Peng, Q.-S.: A global approach for video matching. *Journal of Software* 17(09), 1899–1907 (2006)
11. Wang, W.-L., Tang, L.-L., Li, Q.-Q., Lei, B.: Vehicle detection algorithm with video from airborne camera. *Geomatics and Information Science of Wuhan University* 35(7), 786–789 (2010)

# A Robust Blind Watermarking of Vector Map

Lian-quan Min, Xian-zhen Zhu, and Qiang Li

Institute of Surveying and Mapping, Information Engineering University, Zhengzhou, China  
rainman\_mlq@163.com

**Abstract.** A blind watermarking algorithm is proposed for copyright protection of vector map. The watermarking is embedded by changing the  $x/y$ -coordinates of vertices within the tolerance of the data. The experiments show that the algorithm is fairly good in imperceptibility and robustness against the following attacks: polyline simplifications, moving and cropping of vertices, and addition of small amounts of random noise, renewal operations (adding, deleting, modifying, etc.).

**Keywords:** vector map, watermarking, copyright protection.

## 1 Introduction

Geospatial data has played an irreplaceable role in the social and economic activities. Many businesses, such as, city-planning, environmental protection, vehicle navigation system, project design, resources exploration, land managing, rescue and disaster relief, traveling and military commanding, involve the support of geospatial data. The vector map data is the most important data type in Geospatial information. As digital data, digital vector maps are easy to update, duplicate, and distribute. At the same time, illegal duplication and distribution or forgery of the maps are also easy. Hence, the issue on the protection of their copyrights has been given attention.

First, the acquisition and renewal of vector maps needs large amount of money, time, high precision instruments and so on; therefore, in order to protect the legal rights of the producers and owners of the data, they cannot be used for free. Furthermore, access was restricted for some confidential data, such as military cartographic data. In addition, to protect the user's benefit, the ability of authenticating the source of map data as well as confirming its integrity[1] is required.

Watermarking is an efficient method to protect data from illicit copying and manipulation. Compared with general multimedia data, copyright protection of vector map still drew little attention in the watermarking fields. Unlike general multimedia data, vector map has its distinct features due to the special data structures and application environments of the map data. So the existing methods can not fit the vector map data, and practicability is poor.

Related works will be discussed in Section 2. In Section 3 we will propose a blind watermarking of vector map data. The experimental and the result will be given in Section 4 and Section 5 is the conclusion of the paper.

## 2 The Analysis of Related Works

There are only a few published works on watermarking vector map, and according to the locations where the watermark is embedded, the existing methods for vector map watermarking can be categorized as follows: the algorithms in spatial domain presented by literature [1-11], in transform domain presented by literature[12-21], and some other methods presented by literature[22-26].

Some general issues can be seen by analyzing the current algorithms.

Firstly, the algorithm depends on the order of vertices. The data of a vector map is composed of many coordinates of arranged vertices representing map objects. However, there is no fixed order of vertices in a data file. A scrambled operation on the data affects neither the visual outcome nor the locating precision. It changes the locations of the data storage only. Either reordering the objects in the map, or reordering the vertices within an object can produce a new map file without degrading data's precision. So reordering operation will be a fatal attack for some watermarking scheme which is dependent on the vertices' order.

Secondly, the major consideration on an algorithm is the robustness in resisting the geometrical attacks. The major consideration on most algorithms includes the robustness on resistance of geometric attacks, such as translation, rotation, and zoom. Because the vector map data are normally considered as the basic data for other information system, the authority of the data locating makes clients very cautious in any operation on vector map data; therefore no geometric attack is normally received. And some geometrical transforms, such as translation, rotation, and scaling, change the display only instead of the data themselves. Nevertheless, based on the requirement of the application, vector map data may receive some operations such as cropping, combining, simplification, random noise, as well as renewing operations (adding, deleting, modifying, etc.), which are very common operation to maps data. Therefore, to enhance the practicability of the watermarking system, the algorithm should be immune to cropping, combining, simplification, renewing, random noise, as well as scrambling operations.

Thirdly, watermarking was embedded in map data according to watermarking algorithm of project drawing. We know that project drawing mainly cares about the relative position, while the map data care about not only the relative position, but also the absolute position. Furthermore, the quantum of map data is tremendous compared with the project drawing. So this algorithm has some localization.

Finally, some algorithms partition the whole space into blocks according to rectangle region, which is not universal owing to the unequal distribution of vertices.

## 3 A Blind Watermarking of Vector Map Data Based on Grid

This paper proposes a new robust watermarking of vector map data based on grid. The algorithm can resist cropping, combining, simplification, random noise, as well as renewing attacks such as adding, deleting, modification. Firstly, the watermarking is encrypted to enhance the security; secondly, the map data are processed based on

data grid, and the watermarking is embedded into the map data by modifying the vertices coordinates that have been grid-processed. The watermarking is extracted straightly from the watermarked vector map data, and the original vector map data is not needed, so the algorithm is a blind watermarking, and has significant practical application.

### 3.1 The Vector Map Data Are Processed Based on Data Grid

To enhance the robustness, there is a basic principle that all vertices must be distributed equally into every grid, and the number of vertices in every grid must be no less than the least capacity. If the number of vertices in one grid is too small, the valid watermarking is difficult to detect because its statistical characteristic isn't distinct; and if the number of vertices in one grid is too big, the watermarking is not sensitive to some operations such as modifying, adding, deleting vertices. Hence the rule of classifying vertices is very importance, which is related to the number of all vertices in one vector map and the length of watermarking.

Firstly, we make a 2-dimension chart that has the same size as the watermarking image.

Secondly, read all vertices represented as a pair of coordinates values  $\{(x_i, y_i), i = 1, 2, \dots, N\}$  from map database, in which  $N$  is the number of vertices.

Finally, for every vertex, put it in one grid according to vertex coordinate  $(x_i, y_i)$  and the mapping function:

$$\begin{aligned} S &= f(x_k, x_{k-1} \dots x_0, y_k, y_{k-1} \dots y_0) \\ &= \text{mod}(x_2, x_1, I) + \text{mod}(y_2, y_1, J) \times I \end{aligned}$$

In the function  $I$  and  $J$  are the width and height of watermarking image respectively. Statistically, this process can distribute all vertices equally into grids.

### 3.2 Watermarking Embedding

After step 3.1, all vertices are distributed equally into grids. We embed 1-bit watermarking into one grid. To grid  $A_i$ , we embed  $w_i$  by modifying all vertices coordinates  $(x_i, y_i)$  in this grid according to the special mapping function. In this step, data tolerance  $g$  must be considered to maintain data precision. The detailed steps are described in Table 1.

In table 1,  $nSpecInfo$  and  $nMantissa$  mean the digital in the tens place and the digital in the ones place of  $i$ -vertex coordinate  $(x_i, y_i)$  respectively; A and B mean the parity of  $nSpecInfo$ ; C=4; D=5; and  $g$  is the precision tolerance of vector map data.



**Table 1.** The detailed steps of embedding watermarking

For $i=1$ to $i=N$ ( $N$ is the number of the total vertices )	
Reading coordinate $(x_i, y_i)$ of $i$ -vertex	
Calculating the special information $nSpecInfo$ and the mantissa $nMantissa$ of $i$ -vertex	
Getting 1-bit watermarking $w_j$ that will be embedded into $j$ -Grid where $i$ -vertex lies in	
If ( $w_j = 1$ )	If ( $nSpecInfo = B$ ,and $nMantissa - gw_j < 0$ ),then $vertex[i] = vertex[i] - gw_j$
	If ( $nSpecInfo = B$ ,and $nMantissa + gw_j > 10$ ),then $vertex[i] = vertex[i] + gw_j$
	Else $nMantissa = D$
If ( $w_j = -1$ )	If ( $nSpecInfo = A$ ,and $nMantissa + gw_j < 0$ ),then $vertex[i] = vertex[i] + gw_j$
	If ( $nSpecInfo = A$ ,and $nMantissa - gw_j > 10$ ),then $vertex[i] = vertex[i] - gw_j$
	Else $nMantissa = C$

### 3.3 Watermarking Extracting

While the watermarking is being extracted, the same process based on data grid as embedding is done on all vertices. Next, 1-bit watermarking  $w_j$  is extracted by  $nSpecInfo$  and  $nMantissa$  of vertices in this grid. The watermarking  $w$  can be extracted after getting all  $w_j$ .

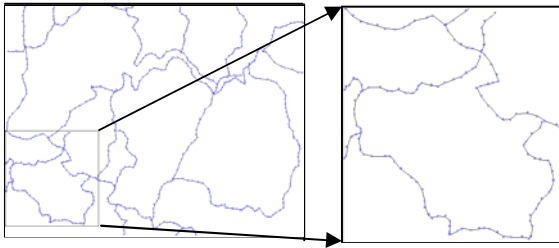
The watermarking is extracted straightly by  $nSpecInfo$  and  $nMantissa$  of vertices in every grid, and the original vector map data are not needed, so the algorithm is a blind watermarking indeed, and has significant practical application.

## 4 The Experiment and the Result

The experimental data are the contour data of a vector map, and have 141311 vertices.

### 4.1 Watermarking Imperceptibility on the Visual Quality of the Map

Fig.1 shows the result for the visual imperceptibility caused by the watermarking embedding. The original contour map is drawn with solid line, and the watermarked contour map is drawn with dash line. The difference of appearance is small enough, and the visual quality can be kept in spite of the embedding. The right figure is an enlarged part of the left one.



**Fig. 1.** The effect of watermarking imperceptibility on the visual quality of the map

### 4.2 Watermarking Imperceptibility on Data Precision

To evaluate watermarking imperceptibility on data precision, the paper takes MSE and Max Error as criterion, the result is shown in table 2.

**Table 2.** The statistics of watermarking imperceptibility on data precision

The error	0	1	2	3	>3
The number of vertices	306162	152661	76877	229911	0
Percent	39.9892	19.9397	10.0412	30.0297	0
MSE	3.30375				

Table 2 shows the geometric error of vertices coordinates caused by embedding. From table 2, we can see that the MSE of vertices coordinates is very small, the max error is less than the tolerance of map data, and the error of near half data is zero. The error caused by watermarking is distributed all over the data space equally and stochastically. So this algorithm is imperceptibility on data precision.

### 4.3 The Resistance against Renewal Operation

Fig.2 shows that the resistance against renewal operation is very good. Fig.2 (a) shows adding 60 percent of the total vertices, (b) shows deleting stochastic 50 percent

of the total vertices, and (c) shows modifying 30 percent of the total vertices. Fig.2 (d) (e) (f) show the watermarking extracted from (a) (b) (c) respectively.

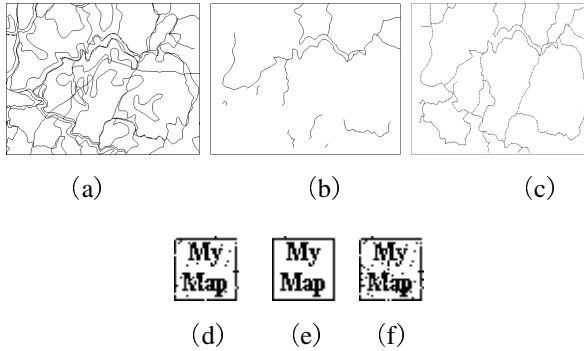


Fig. 2. The resistance against renewal operation

### 4.4 The Resistance against Cropping Operation

In this paper, we use the same cropping patterns as literature[2] to test robustness resiliency to cropping. Each map is cropped according to the 8 cropping patterns shown in Figure 3. Areas of the cropped maps varied from 1/2 to 1/16 of the original.

Fig.3 (i-p) shows the watermarking extracted from (a-h) respectively. From the result, we can see that the algorithm is resistant against cropping operation.

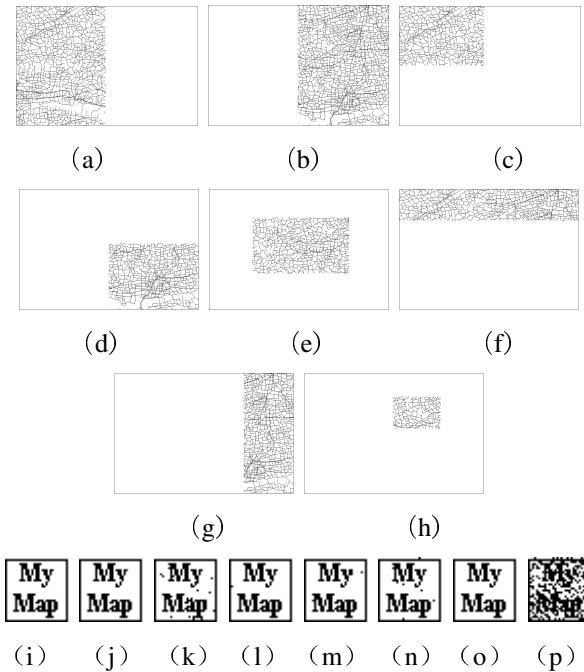


Fig. 3. The resistance against cropping operation

#### 4.5 The Resistance against Additive Random Noise

Fig.4 shows the resistance against additive random noise. The amplitude of the noise in (a) and (b) is 1 and 2 respectively. Fig.4 (c) and (d) are the extracted watermarking from (a) and (b) respectively. The result indicates that the watermarking is resistant against additive random noise if the noise is in tolerance of the vector map data.

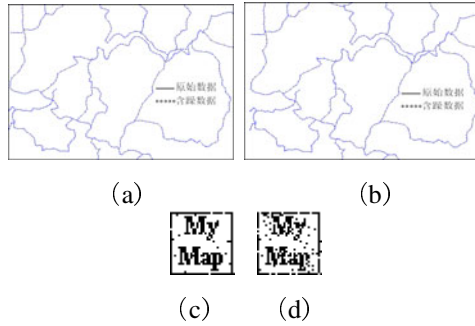


Fig. 4. The resistance against additive random noise

## 5 Conclusions

In this paper, a new digital watermarking algorithm for vector map data is proposed. The experimental evaluation shows that the degradation by watermarking embedding is small, and the robustness is strong for attacks such as polyline simplifications, clipping, noising, renewal operations (adding, deleting, modification, etc.). As a result of these, it is confirmed that the proposed algorithm is effective in protecting the copyright information of the vector map. Future work will include improving the resistance against format transform, the project transform and applying this method to other types of vector data.

**Acknowledgments.** This work was supported by the Nation Natural Science Foundation of China(No. 41071245)

## References

1. Niu, X., Shao, C., Wang, X.T.: A Survey of digital vector map watermarking. Proc. of International Journal of Innovative Computing, Information and Control 2(6), 1301–1316 (2006)
2. Ohbuchi, R., Ueda, H., Endoh, S.: Robust Watermarking of Vector Digital Maps. In: Proc. IEEE International Conference on Multimedia and Expo 2002, Lausanne, Switzerland, August 26-29, vol. 1, pp. 577–580 (2002)
3. Ohbuchi, R., Ueda, H., Endoh, S.: Watermarking 2D vector maps in the mesh-spectral domain. In: Shape Modeling International, Seoul, Korea, pp. 216–228 (2003)

4. Min, L.-Q.: A Robust Digital Watermarking in Cartographic Data in Vector Format. *ACTA Geodaetica et Cartographica Sinica* 37(2), 262–267 (2008)
5. Min, L.-Q.: A Vector Map Data Watermarking Algorithm Based on Statistical Characteristic. *Journal of HAERBIN Institute of Technology* 38(Sup.), 720–724 (2006)
6. Kang, H.: A vector watermarking using the generalized square mask. In: *Proc. of the International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA, pp. 234–236 (2001)
7. Kang, H.I., Kim, K.I., Choe, J.-U.: A Map Data Watermarking Using the Generalized Square Mask. In: *Proc. of the 2001 IEEE International Symposium on Industrial Electronics*, Pusan, Korea, vol. 3, pp. 1956–1958 (2001)
8. Masry, M.A.: A Watermarking Algorithm for Map and Chart Images. In: *Proceedings of the SPIE Conference on Security, Steganography and Watermarking of Multimedia Contents*, vol. VII (January 2005)
9. Voigt, M., Busch, C.: Watermarking 2D-Vector data for geographical information systems. In: *Proc. Of the SPIE, Security and Watermarking of Multimedia Content*, San Jose, USA, vol. 4675, pp. 621–628 (2002)
10. Voigt, M., Busch, C.: Feature-based watermarking of 2D-vector data. In: *Proc. of the SPIE, Security and Watermarking of Multimedia Content*, Santa Clara, USA, vol. 5020, pp. 359–366 (2003)
11. Schulz, G., Voigt, M.: A high capacity watermarking system for digital maps. In: *Proc. of the 2004 Multimedia and Security Workshop on Multimedia and Security*, Magdeburg, Germany, pp. 180–186 (2004)
12. Nikolaidis, N., Pitas, I., Solachidis, V.: Fourier descriptors watermarking of vector graphics images. In: *Proc. of the International Conference on Image Processing*, vol. 3, pp. 10–13 (2000)
13. Solachidis, V., Nikolaidis, N., Pitas, I.: Watermarking polygonal lines using Fourier descriptors. In: *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, vol. IV, pp. 1955–1958 (2000)
14. Kitamura, I., Kanai, S., Kishinami, T.: Copyright protection of vector map using digital watermarking method based on discrete Fourier transform. In: *Proc. of the IEEE 2001 International Symposium on Geoscience and Remote Sensing*, vol. 3, pp. 9–13 (2001)
15. Voigt, M., Yang, B., Busch, C.: Reversible watermarking of 2D-vector data. In: *Proc. of the 2004 Multimedia and Security Workshop on Multimedia and Security*, Magdeburg, Germany, pp. 160–165 (2004)
16. Zhu, C.-Q., Yang, C.-S., Li, Z.-Y.: An Anti-compression Watermarking Algorithm for Vector Map Data. *Journal of Zhengzhou Institute of Surveying and Mapping* 23(4), 281–283 (2006)
17. Li, Y.-Y., Xu, L.-P.: Vector Graphical Objects Watermarking Scheme in Wavelet Domain. *Acta Photonica Sinica* 33(1), 97–100 (2004)
18. Kitamura, I., Kanai, S., Kishinami, T.: Watermarking Vector Digital Map Using Wavelet Transformation. In: *Proceedings of Annual Conference of the Geographical Information Systems Association (GISA)*, Tokyo, Japan, pp. 417–421 (2000)
19. Zhong, S.-P., Gao, Q.-S.: The Feasibility Analysis of Normalized-correlation-based Vector Maps Watermarking Detection Algorithm and the Improved Watermarking Algorithm. *Journal of Image and Graphics* 11(3), 401–409 (2006)
20. Kitamura, I., Kanai, S., Kishinami, T.: Digital watermarking method for vector map based on wavelet transform. In: *Proc. of the Geographic Information Systems Association*, vol. 9, pp. 417–421 (2000)

21. Li, Y., Xu, L.: A blind watermarking of vector graphics images. In: Proc. the Fifth International Conference on Computational Intelligence and Multimedia Applications, pp. 27–30 (2003)
22. Gou, H.-M., Min, W.: Data Hiding in Curves with Application to Fingerprinting Maps. *IEEE Transactions on Signal Processing* 53(10), Part 2, 3988–4005 (2005)
23. Zhou, X., Bi, D.-Y.: Use Digital Watermarking to Protect GIS Data by Chinese Remaindering. *Journal of Image and Graphics* 9(5), 611–615 (2004)
24. Park, K.T., Kim, K.I., Kang, H., Han, S.-S.: Digital geographical map watermarking using polyline interpolation. In: Proc. of the IEEE Pacific Rim Conference on Multimedia, pp. 58–65 (2002)
25. Nikolaidis, N., Pitas, I., Giannoula, A.: Watermarking of sets of polygonal lines using fusion techniques. In: Proc. of the 2002 IEEE International Conference on Multimedia and Expo., vol. 2, pp. 26–29 (2002)
26. Khanna, S., Zaney, F.: Watermarking maps: Hiding information in structured data. In: Proc. of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, USA, pp. 596–605 (2000)

# Implementation of Spectrum Sensing Based on Covariance in Cognitive Radio

Hongyu Li<sup>1,2</sup>, Yuchen Wang<sup>2,3</sup>, Yading Chen<sup>1,2</sup>, and Shaoqian Li<sup>1,2</sup>

<sup>1</sup> Nat. Key Lab. of Sci & Tec on Commun

<sup>2</sup> University of Electronic Science & Technology of China  
Chengdu, P.R. China

<sup>3</sup> School of Electronic Engineering  
{lihy, chenyd, lsq}@uestc.edu.cn

**Abstract.** Spectrum sensing is a fundamental problem in cognitive radio communication system. In this paper, the implementations of spectrum sensing based on the Covariance Absolute Value (CAV) on the test bed of cognitive radio system are presented. This demo cognitive radio system uses the spectrum holes which are determined by sensing for communication with the spectrum from 694MHz to 806MHz and can avoid impact immediately when detecting the primary users (no matter digital television or analog television) at working spectrum. The implementation further reduces the computational complexity and makes a great significance in practical applications. A series of tests show that the detection performance of Covariance Absolute Value is not liable to be affected by the noise uncertainty in practical application and meets the need of the system primly. Furthermore, the performances of detection are also verified with different kinds of source signals.

**Keywords:** CR, Covariance Absolute Value, Spectrum Sensing, Hardware Implementation.

## 1 Introduction

With the fast development of wireless communication and the increasing services in our daily life, the spectrum resources which can be used by free are facing serious challenges in the worldwide range. Cognitive Radio<sup>[1]</sup>, spectrum sensing algorithms are used to detect spectrum holes for communication in this system, which greatly improve the utilization of spectrum resources, is becoming an important direction in the development of wireless communication. At present, many scholars of spectrum sensing mainly focused on theoretical analysis and performance simulation of the energy detection, matched filter detection, covariance detection, cyclostationary feature detection and other methods<sup>[2]-[4]</sup>. These algorithms have their own advantages and disadvantages. For example, the simple energy detection performs well, but it is very sensitive to noise uncertainty. Matched filter detection needs to know a priori information of the test signals, which is often not readily available in the actual environment. Cyclostationary feature detection calls for more hardware resources in the realization because of the high computational complexity.

To gain further knowledge with the performance in the practical application of cognitive radio wireless communication system, we design and build a test platform for dynamic spectrum sharing. Spectrum sensing algorithm of the Covariance Absolute Value was implemented and the performance was also tested at this platform. Tests show that this method can solve the problem of noise uncertainty in energy detection and doesn't need a priori knowledge of the test signal. To meet the detection performance of the system, this method is under the conditions of relatively low computational complexity and easily for implementation with hardware circuit.

The rest of the paper is organized as follows. The hardware and pre-processing of the received signal for sensing are presented in Section 2. Section 3 describes the algorithm of the Covariance Absolute Value and the implementation of it at the hardware platform. The spectrum sensing performances through testing the platform are given in Section 4. Finally, conclusions with the implementation of CAV are drawn in Section 5.

## 2 Pre-processing of Signal

In the dynamic spectrum sharing test platform, we designed eight network nodes, which constitute the cognitive wireless communication system together. The system will co-exist with the existing broadband wireless communication network and be verified at the frequency band from 694MHz to 806MHz which are distributed to television users. The eight network nodes exactly have the same circuit elements, along with the functions of spectrum sensing and information transmission.

### Hardware for Sensing

In the dynamic spectrum sharing test platform, each of the network node is composed of antenna, radio unit, digital signal processing unit and the protocol handler. Among them, the radio unit is made up of low-noise amplifier, frequency synthesizer, filter attenuator, mixer and other components. With 8MHz processing bandwidth, the dynamic range of received signal in radio unit can reach 90dB.

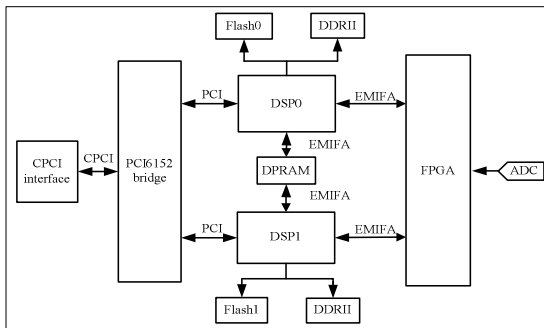


Fig. 1. Block diagram of the digital signal processor board



Digital signal processing unit is shown in Figure 1. It mainly consists of analog-digital conversion chip (ads5485), DSP (TMS320C6455@1G), FPGA (XC5VSX95T), PCI bridge chip and several other memories. The protocol handler will carry out to control the digital signal processing unit through the standard bus interface CPCI when sensing.

### Pre-processing of Sensing

Electromagnetic signals will be received by coupling to the RF unit through the antenna. After down conversion, low-noise amplifier, filters and amplification, the received signal could be converted to low intermediate frequency signal with the center frequency of 25MHz and 8MHz bandwidth which will be sent to the 16-bit analog-digital conversion circuit. The signals after the A/D processing will be processed through I/Q demodulation and filter again. Automatic Gain Control (AGC) module measures the input signal and manages the Gain Controller in RF unit based on the measurement results all the time. The stability data after AGC will be stored into DSP memory according to system parameters for sensing operation. I/Q modulation, filter and AGC are implemented in FPGA. The main pre-processing procedures of signal are shown below.

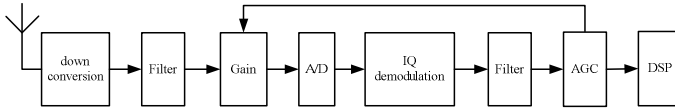


Fig. 2. Pre-processing of sensing signal

## 3 Spectrum Sensing Implementation

In 2007, Yonghong Zeng, in Singapore Telecommunication Research Institute, made a variety of sensing algorithms based on the signal characteristics of covariance<sup>[5]-[8]</sup>, such as the Covariance Absolute Value (CAV), Covariance Frobenius Norm (CFN), and Maximum-Minimum Eigenvalue (MME). These algorithms have a great significance for the study of spectrum sensing in Cognitive Radio. The implementation process of covariance detection is shown in Figure 3.

As described above, a variety of data processing methods can be used after obtaining the covariance. The covariance Absolute Value (CAV) chosen in hardware implementation at our system will be presented below.

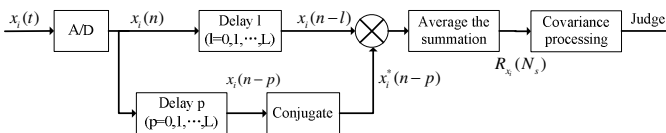


Fig. 3. Block diagram of covariance detection algorithm

**Covariance Absolute Value (CAV)**

Assume that  $x(n)$  is the received signal, which are filtered by filter and converted by A/D. The sample auto-correlation of the received signal can be expressed as:

$$\lambda(l) = \frac{1}{N_s} \sum_{m=0}^{N_s-1} \tilde{x}(m)\tilde{x}^*(m-l), l=0,1,\dots,L-1 \tag{1}$$

where  $N_s$  is the length of the received signal and  $L$  is called smoothing factor.

Then, the statistical covariance matrix of the received signal can be approximated by the sample covariance matrix. It is defined as:

$$R_x(N_s) = \begin{bmatrix} \lambda(0) & \lambda(1) & \dots & \lambda(L-1) \\ \lambda^*(1) & \lambda(0) & \dots & \lambda(L-2) \\ \dots & \dots & \dots & \dots \\ \lambda^*(L-1) & \lambda^*(L-2) & \dots & \lambda(0) \end{bmatrix} \tag{2}$$

Note that the sample covariance matrix is Hermitian and Toeplitz. From the statistical covariance matrix, we can see that the non-diagonal elements of the matrix are zero if the signal does not exist. On the other hand, when the signal is present and there is a correlation between the signal sampling (due to over-sampling, multipath effects at transmission, correlation of the original signal and other reasons, the sampling signal is correlative at receiver in general), the non-diagonal elements of the matrix exist nonzero values. Therefore, define:

$$T_1(N_s) = \frac{1}{L} \sum_{n=1}^L \sum_{m=1}^L |r_{nm}(N_s)| \tag{3}$$

$$T_2(N_s) = \frac{1}{L} \sum_{n=1}^L |r_{nn}(N_s)| \tag{4}$$

where  $r_{nm}(N_s)$  is the element at n row m column of the matrix  $R_x(N_s)$ . It can be see: when the signal does not exist,  $T_1 = T_2$ ; when the signal is present,  $T_1 > T_2$ . Hence, we can detect the signal existence by comparing  $T_1$  and  $T_2$ .

The detection statistic of CAV is:

$$T = T_1(N_s) / T_2(N_s) \tag{5}$$

When the signal does not exist, from literature [6] we can deduce that:

$$\begin{aligned} T &= T_1(N_s) / T_2(N_s) \approx E(T_1(N_s)) / E(T_2(N_s)) \\ &= 1 + (L-1)\sqrt{2 / (\pi N_s)} \end{aligned} \tag{6}$$

And the threshold can be expressed as:

$$\gamma = \frac{1 + (L-1)\sqrt{2 / (\pi N_s)}}{1 - Q^{-1}(P_f)\sqrt{2 / N_s}} \tag{7}$$

where  $P_f$  is the probability of false alarm, and

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-u^2/2} du \quad (8)$$

It can be seen from the formulas (6) and (7) that the detection statistic and threshold are nothing to do with noise power when the signal is not present. That is, no matter how the noise power changes in environment, the detection statistic and threshold are not affected, so the detection performance was not affected, too. Therefore, CAV can be a good solution in the presence of uncertainty problem in energy detection and with low computational complexity.

### Implementation Framework

Sensing channel in the system mainly include DSP and FPGA. FPGA is responsible for pre-processing the signal which is from RF unit and DSP complete the work of spectrum sensing with the input data. MAC is the protocol handlers which control the physical layer and make networking in the system.

The sensing data transmission between DSP and FPGA is through the interface of EMIFA. In this test platform, EMIFA was configured as synchronous and 16-bit. PCI play a role of mapping which made easily to transfer commands between DSP and MAC.

The network nodes in dynamic spectrum sharing test platform work as secondary users on the spectrum from 694MHz to 806MHz with the primary users of television. In the system, when a signal is detected, we need to further determine it is a primary user or a secondary user. So, the active period and silence period are designed to distinguish them.

In active time, the network can communicate with each other. But in silence time, all of the nodes in the network should be the status of no-communication. The primary users of television can take the spectrum arbitrarily at any time. This TDD mode can help the network nodes to distinguish the primary users and secondary users.

### CAV Implementation

From the formula (1), we can get the sample auto-correlation of the received signal as:

$$\begin{aligned} \lambda(l) &= \frac{1}{N_s} \sum_{m=0}^{N_s-1} \tilde{x}(m) \tilde{x}^*(m-l) \\ &= \frac{1}{N_s} \sum_{m=0}^{N_s-1} [a(m) + jb(m)][a(m-l) - jb(m-l)] \\ &= \frac{1}{N_s} \sum_{m=0}^{N_s-1} ([a(m) \times a(m-l) + b(m) \times b(m-l)] + \\ &\quad j[-a(m) \times b(m-l) + b(m) \times a(m-l)]) \end{aligned} \quad (9)$$

we assume that  $\tilde{x}(m) = a(m) + jb(m)$ .

To get the covariance matrix of signal, we only need to calculate the value from  $\lambda(0)$  to  $\lambda(L-1)$  because the matrix is Hermitian and Toeplitz which have been described above.

When implementing, the sensing data stored in DSP memory with real part  $a(m)$  and imaginary part  $b(m)$  will be converted into a 2-byte word.

The processes of implementation of CAV are as follows:

Step1: calculate the real part of  $\lambda$

$$\lambda\_real(l) = \sum_{m=0}^{N_s-1} [a(m) \times a(m-l) + b(m) \times b(m-l)] \quad (10)$$

Step2: calculate the imaginary part of  $\lambda$

$$\lambda\_imag(l) = \sum_{m=0}^{N_s-1} [-a(m) \times b(m-l) + b(m) \times a(m-l)] \quad (11)$$

Step3: 10 bits right shift of  $\lambda\_real(l)$  and  $\lambda\_imag(l)$  to void over-range of the results

Step4: calculate the amplitude of  $\lambda$

$$\lambda(l) = \lambda\_real(l) * \lambda\_real(l) + \lambda\_imag(l) * \lambda\_imag(l) \quad (12)$$

Step5: calculate  $T_1(N_s)$  and  $T_2(N_s)$

$$T_1(N_s) = \sum_{l=0}^L \lambda(l) \quad T_2(N_s) = \lambda(0) \quad (13)$$

Step6: calculate  $T = T_1(N_s) / T_2(N_s)$  and compare with threshold  $\gamma$  for detection

The computational complexity of CAV implementation are mainly at step1 and step2 for calculating the real and imaginary part of  $\lambda$ . In hardware, we can segment the received data skillfully for parallel operation to improve the response time of system. Through converting complex multiplication to real multiplication, the total computational complexity of CAV only needs  $4N_sL + 2L$  real multiplications and  $2(N_s + 1)L$  additions. It reduces computing operations a lot than theoretical analysis<sup>[5]</sup>. The threshold  $\gamma$  is determined by test with the fixed probability of false alarm  $P_f$  in practically.

## 4 Performance Test

In the following, we will give some test results using the digital video DTMB and PAL standard analog TV signal. DTMB signal can be produced by Digital Video of Agilent Signal Studio.

The DTMB signals were down converted to a low central IF frequency of 25MHz and sampling at 100M samples/sec. The analog-to-digital conversion of the RF signal used a 16-bit A/D. Each sample was encoded into a 2-byte word (signed int16 with a two's complement format). Before be sent to DSP for sensing, ten times of under

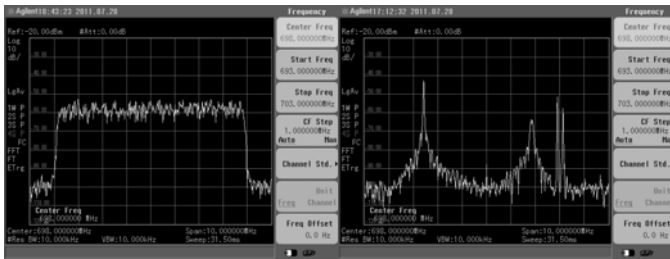
sampling will be operated. In order to reduce the computational complexity of sensing, the number of used samples is 4096 and 8192 (corresponding to 0.4ms and 0.8ms), and the smoothing factor is chosen as  $L=10$ . The threshold is set based on the  $P_f = 0.1$  and fixed for signals.

The settings of DTMB signal in testing are shown below in table 1.

**Table 1.** The settings of DTMB signal

Modulation Type	16 QAM	Symbol Rate	7.56MHz
Interleave Mode	B=52,M=720	Code Rate	0.4
Sub-Carrier Number	3780	Number of Super Frame	0.05
Filter Type	Root Nyquist	Filter Length	64
Roll-off Factor	0.05		

We also test the PAL standard analog TV signal. This signal was captured in real environment and stored to Agilent Signal Source by ftp. The TV signals set with sample rate at 10.24MHz and offset -2.58MHz to pilot.



**Fig. 4.** Spectrum of DTMB and TV signal (PAL)

Figure 5 gives the probability of detection results based on the DTMB signal with different signal amplitude. Figure 6 gives the results based on the PAL standard TV signal. The detection probability at each point in the figures with the amplitude of signal is obtained through the test of 10000 times.

Tests show that the threshold will not be affected by the noise power in environment. The larger number of samples used for sensing, the better detection performance will be brought, and this feature has been particularly evident in Figure 6. The detection time of 0.8ms outperforms the 0.4ms sensing with a gain of about 5-6dB under the condition of PAL standard analog TV signal. Generally, the detection performance of PAL TV signal is better than DTMB signal.

In summary, the tests show that the CAV method of sensing works well without using information of the original signal and noise power. The detection performance can meet the needs of the cognitive radio system.

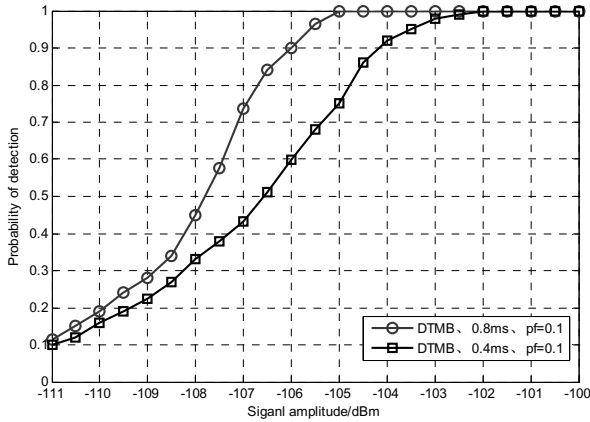


Fig. 5. Performance of detection with DTMB signal

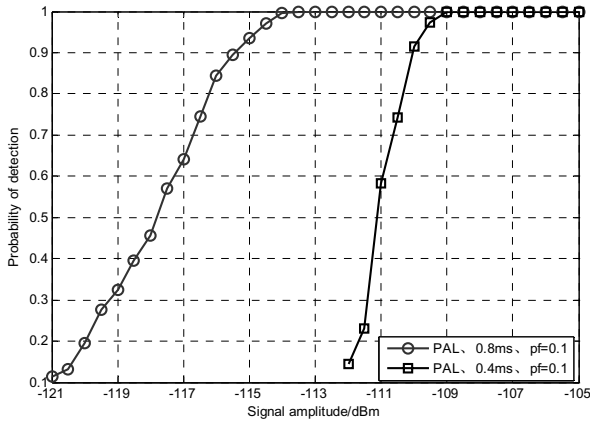


Fig. 6. Performance of detection with PAL signal

### 5 Conclusion

On the cognitive radio system working with primary users of television on the spectrum from 694MHz to 806MHz, chosen the reasonable sensing algorithm based on covariance, we implement the spectrum sensing on this test bed. Through the aborative design, the process of algorithm implementation further reduces the computational complexity. Both the DTMB and PAL standard TV signal have been verified with this method. Tests show that the threshold of detection is independent with the noise power, only be relative to the length of processing data. The detection performance of analog TV signal is superior to the digital television signal.

**Acknowledgments.** This work is supported in part by the High-Tech Research and Development Program (863 Program) of China under Grant No.2009AA011801 and 2009AA012002, and National Basic Research Program (973 Program) of China under Grant No.2009CB320405, National Grand Special Science and Technology Project of China under Grant No.2009ZX03007-004, 2009ZX03005-002, 2009ZX03005-004,2010ZX03006-002-02.

## References

1. Mitola, M.G.: Cognitive radio: making software radios more personal. *IEEE Personal Communications Magazine* 6(4), 13–18 (1996)
2. Sahai, A., et al.: Fundamental tradeoffs for opportunistic radio-systems. In: *Workshop of Cognitive Radio, Software Defined Radio and Adaptive Wireless System (RWS 2006)* (January 2006)
3. Cabric, D., et al.: Implementation issues in spectrum sensing cognitive radios. In: *Asilomar Conference on Signal, Systems and Computers* (November 2004)
4. Gardner, W.A.: Signal Interception: A Unifying Theoretical Framework for Feature Detection. *IEEE Transactions on Communication* 36(8) (August 1988)
5. Zeng, Y., Liang, Y.: Covariance Based Signal Detections for Cognitive Radio. In: *Proc. IEEE in Dynamic Spectrum Access Networks*, pp. 202–207 (2007)
6. Zeng, Y., Liang, Y.: Spectrum sensing algorithms for cognitive radio based on statistical covariances. *IEEE Transactions on Vehicular Technology* 58(4), 1804–1815 (2009)
7. Zeng, Y., Koh, C.L., Liang, Y.-C.: Maximum Eigenvalue Detection: Theory and Application. In: *IEEE International Conference on Communications*, pp. 4160–4164 (2008)
8. Zeng, Y., Liang, Y.-C.: Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Transactions on Communications* 57(6), 1784–1793 (2009)

# Simulation and Analysis of Quiet DDoS Attacks

Jing Zhang, Bo Liu, Huaping Hu, Lin Chen, and Tianzuo Wang

School of Computer, National University of Defense Technology  
Changsha, China

jingzhang132@gmail.com, boliu615@yahoo.com.cn,  
howardnudt@yahoo.com.cn, leeyon@yahoo.cn, phoenixwtz@163.com

**Abstract.** The Quiet DDoS attack was proposed in 2009 by Amey and Nirwan. It launches certain number of TCP flows to reduce the transferring ability of normal TCP flows by using botnet. By the simulation experiments on the NS2 platform, we analyze the factors and their impacts on the attack in detail. According to the results of simulation, we give some suggestion to counter the Quiet DDoS attack.

**Keywords:** quiet, DDoS, attack, simulation, counter.

## 1 Introduction

Nowadays, the society increasingly depends on the network, for instance, e-business, e-government, e-bank, and so on. Since the network has certain weakness, it faces with many threats. The Denial-of-Service (DoS) attack is one of these threats with increasing serious threat to the internet. The goal of the DoS attack is to prevent legal user to get service or resource from the internet. Using the connection with the internet, through the malicious operation on the victim end system or mass useless data transferred on the internet, the attackers achieves their goal. Because of the number of node which participates in the attack become larger, the DoS attack is called Distributed DoS (DDoS) attack which is frequently used to launch the attack.

There are two reasons to make the DoS attack have so severe threat to the network safety and its application. One reason is the defend mechanism cannot detect the attack effectively. The other reason is the attack flow has higher stealthy and legal. For example, the LDoS attack which proposed in 2003 [1], and followed by RoQ attack [2-3] and Pulsing Attacks [4], make the steal of the attack flow have greatly improved. The Quiet DDoS attack [5] adopt the TCP protocol to generate the attack flow, not only making the attack legalize in the appearance, but also has higher stealthy because of the low attack traffic rate. The characteristic of the attack flow make the attack have severe threat to the network. So, the Quiet DDoS attack is the focal research of the paper. How the degree of threat which the attack have, and the possible counter technologies will be discussed in the paper.



## 2 The Quiet DDOS Attack

The research on the interaction among normal TCP flows has been done in [6] by Shirin etc. The principle of the Quiet DDoS attack uses the short-lived TCP flow impact on other TCP flows which is similar with the result which is got in [6]. The execution phase of the attack has three steps, showed as follows [5]:

1. The attacker issuing a command to the bots to trace the path from the bot to the certain website, the router whose IP address is observed in all the trace outputs becomes the target for attack by the attacker, suppose the target router is A.
2. Then the attacker collects the set of web servers which transfers the data by A, suppose the set is S.
3. Launching the attack, using the attack strategy, which is showered as Fig.1, make the attack flow distributed in S.

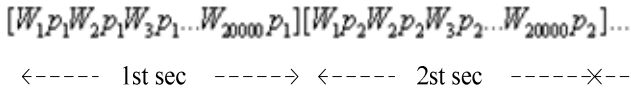


Fig. 1. Attack Strategy

From the introduction of the Quiet DDoS attack, we can get the two characteristics of the attack which is showered as follows.

1. The attack flow is the TCP flow which is consistent with the normal TCP flow in the internet. Compared with the LDoS attack, the attack flow has higher stealthy.
2. The attack flow no longer concentrating on few IP addresses, but distributed in appearance which makes the algorithm to counter the attack by analyzing the target IP address fail to counter the attack.

From the above introduction, we can know that the Quiet DDoS attack has higher stealthy and fails to be countered by the mechanism deployed in the internet, so the attack will bring serious threat to the network safety. In order to quantitative analysis the threat, we give the metric in the paper.

Basically speaking, the attacker can modify the parameters of the attack, includes attack period, attack flow, the size of attack file, and the size of botnet. And also the diversification of the network, includes different link delay and number of hop from source to the destination, may impact the attack performance. How these factors impact on the attack performance will be discussed in the paper. The paper implements the Quiet DDoS attack by simulation to analyze the impaction. Through simulation, we give the impaction of the above related factors make on the attack, improve or reduce. According to the corresponding conclusion, we can give some suggestions to counter the Quiet DDoS attack.

### 3 Simulations

#### Settings

In this paper, we use NS2 [7] simulation environment to build simulation network [5] based on the Fig. 2.  $C_0$ ,  $C_1$  generate normal network traffic, including five FTP flows and the traffic of HTTP that generate by the PackMime module, which occupy 30% of the bottleneck link bandwidth. The normal flows that represent the long-lived TCP flow which will be affected by the attack are generated by  $N_1, \dots, N_5$ . The attack flow generates by  $S_1, \dots, S_{10}$ . In an attack period, each attack node sends a file with the size of 750KB, while the node  $AD_1, \dots, AD_{10}$  receives the attack flow. The node  $A_{11}, \dots, A_N$  with corresponding receives node is also the attack node to analyze the size of the botnet impacts on the attack performance in the following section. In order to match the actual network circumstance, we make some nodes join in the network. The queue management deployed in the network is Droptail.

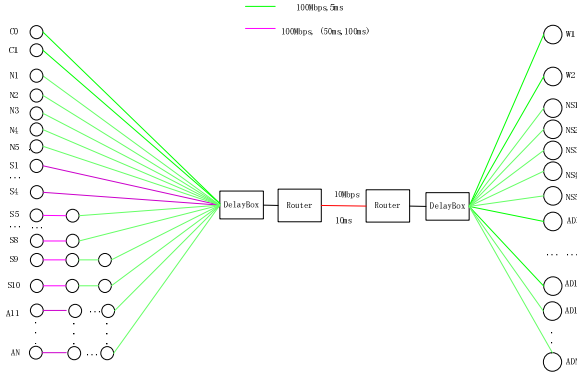


Fig. 2. Network Topology

#### Metrics

Under normal circumstances, the bytes of transferred data which are totally sent by normal flow are  $DataTransfer\_Normal$ . Under attacking, the bytes are  $DataTransfer\_Attack$ . The relationship between the two values is  $DataTransfer\_Normal \geq DataTransfer\_Attack$ . So we have the following definition of *Reduction* as the metrics to weigh the attack performance of the Quiet DDoS attack.

Definition1:

$$Reduction = \frac{DataTransfer\_Normal - DataTransfer\_Attack}{DataTransfer\_Normal} \quad (1)$$

In this paper, we use this metric to estimate the attack performance of the Quiet DDoS attack. As can be seen from the definition, the best attack performance is 1. In this case, the affected long-lived TCP flow cannot get any services. The worst attack performance is 0; in the best case, the attack does not have any effect on the long-lived TCP flow.

## 4 Simulation Results and Analysis

In this chapter, we mainly discuss the factor which impacts the attack performance. The number of node to send the attack flow is 10 except the section of “Size of Botnet”.

### Attack Period

In the literature [5], the attack period is randomly distributed in  $(0, 1s]$ . Whether the attack performance will be the best under the maximum value of the attack period set as 1s, we will discuss the setting of the attack period in this section. In the simulation, we make the attack period with the fixed value to better analyze the attack.

With the same setting in simulation with [5], we make the attack period distributed in  $[0.5s, 100s]$  to analyze the changes of the attack performance. The attack performance with different attack period is showered as Fig.3. We can get the conclusion that the attack performance is not becomes better if the attack period becomes small. The optimum attack period is 6s.

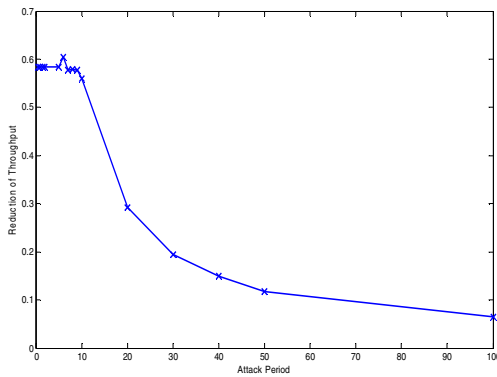


Fig. 3. Attack Period

From the setting of the simulation, we know that, in an attack period, each node sends 750KB attack flow, the total attack flow is 60Mb, so the average attack rate is 10Mbps which is equal with the capacity of the bottleneck link. Suppose that: a) *File\_Size* denotes the attack flow in an attack period, b) *Num\_Attack* denotes the number of attack node or the size of botnet, and c) *Attack\_Period* denotes the

attack period, d)  $C$  denotes the capacity of the bottleneck link. So the optimum attack period can be calculated as follows:

$$File\_Size * Num\_Attack / Attack\_Period = C \quad (2)$$

$$\Rightarrow Attack\_Period = File\_Size * Num\_Attack / C \quad (3)$$

In order to validate the accuracy of the formula, while the number of the attack node is fixed with 10, through adjusting the value of the  $File\_Size$ , we get the attack performance under different attack period which get by calculating the attack period through the formula (3). The data is showered in Tab.1. From the data in Tab 1 we can know that the attack performance has small difference with different attack period, so we have the conclusion that the formula (3) is accurate to determine the attack period.

**Table 1.** Attack Period

File Size	Attack Period	Reduction
5000KB	40s	0.6015
4500KB	36s	0.5845
3750KB	30s	0.5845
3000KB	24s	0.5845
2250KB	18s	0.5766
1500KB	12s	0.6042
750KB	6s	0.6042
375KB	3s	0.6042
250KB	2s	0.5811
125KB	1s	0.6009

### Attack Flow

In this section, we will discuss different attack flow have how impacts on the attack performance. The simulation is done in two values of attack period, 6s which got in the above section and 1s which used in the literature [5].

From the setting of the simulation network topology we know that the number of attack node is 10 with different linkdelay. How the attack performance is if only one node to launch the attack, the data is showered as Fig.4. In the same attack period, different node impacts on the attack performance has small difference. With the smaller attack period, the attack performance is better which consistent with the conclusion in [6].

If only parts of nodes participating in the attack, the total attack traffic will reduce, the changes of the attack performance are showered as Fig.5. Along with the total attack flow increases, the attack performance becomes better.

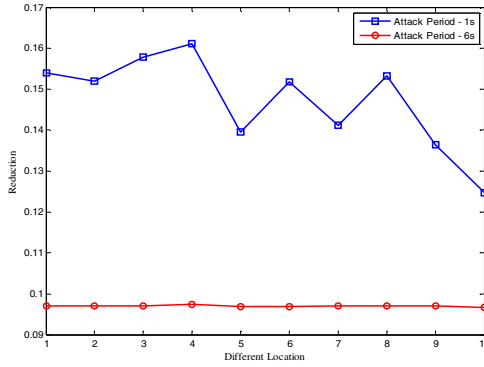


Fig. 4. Single Node

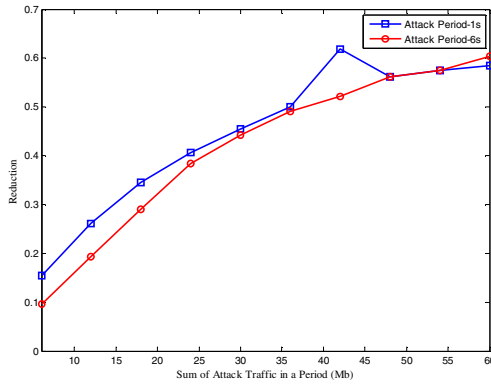


Fig. 5. Attack Flow

### Linkdelay

In the above simulations, the linkdelay does not have any change. If we increase the linkdelay from the attack node to the node that receives the attack flow, how changes the attack performance have, we will discuss in this section. The simulation results are showered as Tab.2 and Tab.3. From the data we know that the attack performance of the Quiet DDoS attack has small changes with different linkdelay, the attack has lower request to the network environment.

**Table 2.** Add Fixed Link Delay

<b>Add Fixed Linkdelay</b>	<b>Reduction</b>	<b>Add Fixed Linkdelay</b>	<b>Reduction</b>
0	0.6042	30ms	0.5776
10ms	0.5838	40ms	0.5685
15ms	0.5871	50ms	0.5489
20ms	0.5830	100ms	0.5251
25ms	0.5684	200ms	0.4554

**Table 3.** Add Random Link Delay

<b>Add Random Link delay</b>	<b>Reduction</b>	<b>Add Random Link delay</b>	<b>Reduction</b>
(0, 10ms)	0.5823	(20ms, 80ms)	0.5572
(10ms, 70ms)	0.5610	(30ms, 70ms)	0.5441
(20ms, 70ms)	0.5370	(30ms, 60ms)	0.5520

The above simulation only increasing the linkdelay, if we add the hop between the attack node and the receive node, whether the attack performance still good enough to reduce the capacity of the data transfer for the normal flow, we will discuss here. We make some simulations to validate that, the data is showered as Tab.4. From the simulation result we can know that the attack performance is still good.

**Table 4.** Add Fixed Hop

<b>Number of Hop</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>Reduction</b>	0.6042	0.5989	0.6201	0.6165	0.6165

Through the above simulation, the conclusion we get is that the Quiet DDoS attack strongly robust to the network environment.

### Size of Attack File

In the literature [5], each node sends the same attack flow in each period. If the node sends different attack flow in a period while keeping the total attack flow unchanged, how change the attack performance have, the result is showered as Tab.5. From the data of Tab.5, we can know, if the attacker suitably assigns the total attack flow, the attack performance with the false random attack flow is good enough which will cover-up the characteristics.

**Table 5.** Simulation Result

Different File Size - Number		Reduction
1200KB-1	700KB-9	0.5964
1350KB-2	600KB-8	0.5920
1125KB-4	500KB-6	0.5898
700KB-5	800KB-5	0.5964
500KB-3	860KB-7	0.5904
400KB-5	1100KB-5	0.5649
300KB-6	1425KB-4	0.5072
200KB-7	2034KB-3	0.4316

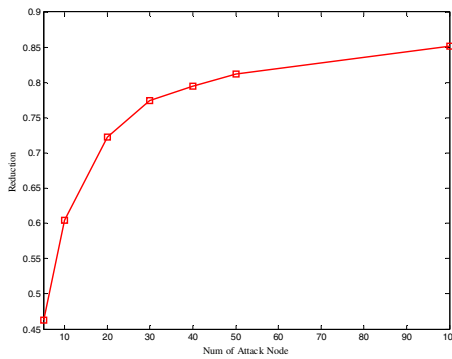
**Size of Botnet**

If the size of the botnet changes, larger or smaller, how changes the attack performance has will be discussed in this section.

The symbols,  $C$ ,  $File\_Size$ ,  $Num\_Attack$  and  $Attack\_Period$ , has the same meanings as the section of “attack period”. Then the attack flow of each node sends in a period can be computed as follows:

$$(C * Attack\_Period)/(Num\_Attack * 8) \tag{4}$$

The unit is Bytes. The other settings are same with the above simulations. The attack period is 6s based on the above conclusion. The number of attack node impacts on the attack performance is showered as Fig.6.



**Fig. 6.** Number of Attack Node

From the Fig.6 we can see that, along with increasing the number of the attack node, the attack performance becomes better. Take two attacks as examples, suppose the number of the attack node is  $M_1$  and  $M_2$  respectively, the relationship between them is  $M_1 < M_2$ . In these two attacks, the first packet to be dropped is  $t_1$  and  $t_2$  respectively. Sample the number of packet transferred by the router using fixed

sample time during the attack time, so we get the two sequences for the number of the packet sending by the normal flow, they are  $n'_1, n'_2, \dots, n'_{t_2}$  and  $n_1, n_2, \dots, n_{t_2}$  respectively. The sequences for the attack flow are  $M_1, m'_1, m'_2, \dots, m'_{t_2}$  and  $M_2, m_1, m_2, \dots, m_{t_2}$ . Because the time  $t_1$  and  $t_2$  is the time to drop first packet, all the flow do not affect by the retransmission timeout or the three same ACK packet before that time. So, setting time  $j$  that satisfies  $j \leq \min(t_1, t_2)$ , combines with the preconditions, we can know that  $m'_i < m_i, n'_i < n_i (i \leq j)$ . The conclusion we can get from the above analysis is follow.

**Conclusion 1.** Attack the same router, the time to drop first packet satisfies  $t_1 > t_2$ .

By the introduction of the front chapter, we can know that the attack flow cannot be all transferred in a period. And we also get the second conclusion.

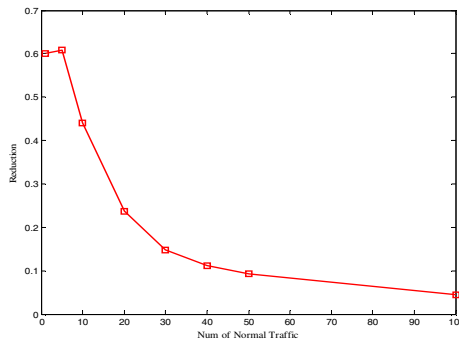
**Conclusion 2.** Partly of the attack flow will be add up to the next period.

The lager number of the attack node, the more time will the attack flow impacts on the normal flow, the better performance.

Two above two conclusions are very good to explain the reason to cause the attack performance of the Quiet DDoS attack with the increment number of the attack node.

### Number of Normal TCP Flows

In this section, we will discuss the increment number of the normal flow have how impacts on the attack performance. The attack period still chooses 6s, and the attack node is 10. The result is showered as Fig.7. As shown by Fig.7, along with the increment number of the normal flow, the attack performance becomes worsen. At this time, change the role of the normal flow and the attack flow, using the theory in "Size of Botnet" section, we can explain the reason to cause this result.



**Fig. 7.** Number of Normal Flow



In the simulation data, we find that, the capacity of some normal flows will be raise although the attack will reduce the data transferred capacity by the most part of the normal flow along with the increment number of normal flow. Make the 100 normal flows as the example. The capacity of data transferred for each node in the simulation is showered as Fig.8. Compared attack and normal two instances, the total capacity of transferring data still reduce under attack, but the capacity of partly normal flows will increase.

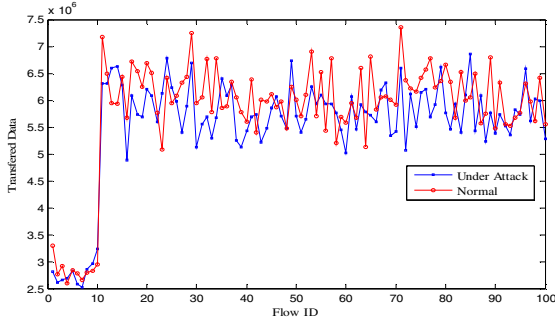


Fig. 8. Compared the Two Instances

## 5 Counter Technologies

From the above introduction and analysis of the Quiet DDoS attack, we summarize the attack have following characteristics.

- 1) Suit distributed attack very much.
- 2) Adopt completely legal flow in appearance, no longer limited on the UDP flow.
- 3) The attack performance is good.
- 4) The attack period can be selected a bigger value.

From the four characteristics, we can see that the Quiet DDoS attack make the attack flow have higher stealthy and legal. And also, small changes on the attack parameters, such as the size of botnet, can improve the attack performance. So, this attack will be a serious threat to the network safety.

According to the analysis of the factors impacts on the attack performance, we find that only improving the number of the normal TCP flow or reducing the attack flow can decrease the attack performance. But in the certain circumstances, the number of the normal TCP flow is fixed, what we can do to mitigate the attack performance is to filter the attack flow. We must distinguish the attack flow from the normal flow to filter it. Although the attack flows completely legal in appearance, it still has weakness, such as short-lived but periodically for the fixed TCP connection. So, the way to filter the attack flow is to extract the signature of the attack flow based on the characteristic of the periodically sending the short-lived attack flow, then to filter the flow which satisfies the signature. There still has another way through establishing the network behavior habit for each node to filter the flow which destroys the habit. These two ways to filter the flow are our future work.

## 6 Conclusion

The paper analyzes the principle of the Quiet DDoS attack, and simulates on the NS2 platform to analyze the factors impacts on the attack performance. We can see that this attack can select big attack period and false randomly attack flow from the conclusion we get. So, the related research on this attack is important. In the end, we give some possibly counter technologies on this attack.

**Acknowledgments.** We thank Amey and Nirwan for providing the source code to me. We would like to thank the High Technology Research and Development Program of China (863 Program) for support of this project with grant 2008AA01Z414.

## References

1. Kuzmanovic, A., Knightly, E.W.: Low-rate TCP-targeted denial-of-service attacks. In: Proceedings of ACM SIGCOMM 2003, Karlsruhe, Germany, pp. 75–86 (2003)
2. Guirguis, M., Bestavros, A., Matta, L.: Exploiting the transients of adaptation for RoQ attacks on Internet resources. In: ICNP 2004, Berlin, Germany, pp. 184–195 (2004)
3. Guirguis, M.: Reduction of Quality (RoQ) Attacks on Internet end-systems. In: Proceedings of the 24th IEEE INFOCOM, Miami, Florida (2005)
4. Luo, X., Chang, R.: On a new class of pulsing denial-of-service attacks and the defense. In: Proceedings of Network and Distributed System Security Symposium (NDSS 2005), San Diego, CA (2005)
5. Shevtekar, A., Ansari, N.: Is It Congestion or a DDoS Attack? *IEEE Communications Letters* 13(7), 546–548 (2009)
6. Ebrahimi-Taghizadeh, S., Helmy, A., Gupta, S.: TCP vs. TCP: a Systematic Study of Adverse Impact of Short-lived TCP Flows on Long-lived TCP Flows. In: Proceedings of the 24th IEEE INFOCOM, Miami, Florida (2005)
7. McCanne, S., Floyd, S.: *The network simulator-ns-2* (2010)

# VV&A Resources Management System of Guidance Simulation

Zhenjun Li and Peng Jiao

National University of Defense Technology, Changsha 410073, China

**Abstract.** In order to heighten the automation degree of VV&A and the security of resource information, the effective management of VV&A resource is needed. In this paper, we firstly analyze the basic flow of lifecycle VV&A of guidance simulation system (GSS) and the related resources during every phase of VV&A. And then the model of VV&A resource will be established. In order to get effective management of VV&A resource, database of VV&A resource is developed based upon Oracle 10g DBMS. Finally VV&A resource management system is designed and realized.

**Keywords:** Guidance Simulation, VV&A, Database, Resource Modeling, Resource Management.

## 1 Introduction

In order to implement VV&A tasks successfully, lots of resources are needed, such as test data, experiment data, different categories of documents, pictures and videos. Therefore, a software tool named VV&A Resources Management SYSTEM (VRMS) is designed and developed to support VV&A tasks of GSS. The Structure and functions of VRMS are shown as follows.

## 2 Categories of VV&A Resources

VV&A resources are produced mainly in the development process of GSS, and the development process of GSS is depicted in figure 1.

As shown in Fig.1, in the “Requirements Phase”, the corresponding VV&A tasks are verification of simulation requirements and generating verification reports, hence the resources in this phase are various requirement documents. In the “Design Phase”, the corresponding VV&A tasks are verifying mathematical models, components scheme, and design reports etc. As a result, the relative resources in this phase mainly include kinds of design specifications or schemes of various models, and the simulation programs to validate the validity of mathematical model are important as well. At last in the “Implement Phase”, the mainly resources are various kinds of test data, experiment data, and relative analysis reports. In addition, for validating the validity of simulation experiment data and simulation model, flight experiment data are also needed. And in order to review simulation experiment process and operate the model, some pictures and videos should be gotten as Fig.1.

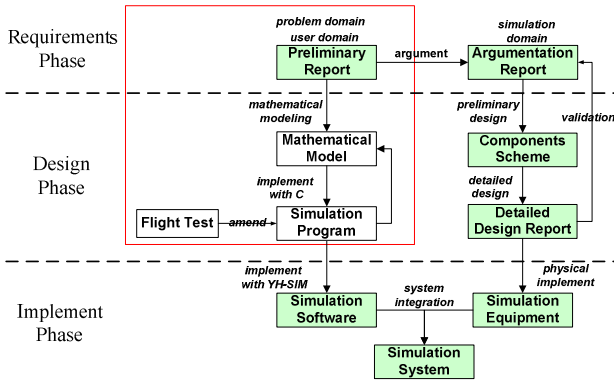


Fig. 1. The development process of GSS

In conclusion, VV&A resources are the set of all valuable products in the whole life-cycle of GSS, which include data, documents, models, pictures, videos and so on.

### 3 Resources Modeling

There are three reasons for resources modeling: a) Find valuable information from properties of various resources, and establish the data model of it, b) Provide the foundation for data structure and management rules of bottom database, c) Provide the model formwork for different guidance simulation system and improve the reusability of model.

In this paper, the IDEF1X method is used to describe the models of document, mathematical model, data, video and other resources based on the software tool – “ERWin Data Modeler”. The logical model should be gotten firstly, and then transformed to a whole physical model.

#### Logical Model

Each kind of resource has some different properties, as to document, the properties of which include document title, generating date, author (or organization), abstract, purpose and so on. But not all the properties are concerned by VV&A practicers. Besides, since the sources of different resources are the same with each other, the properties of which must also be same as others or relative with others. Logical Model is a kind of information model which provides a specification to describe the entity and property of resources.

For example, as for mathematical simulation experiment result data, the logical model is depicted in figure 2.

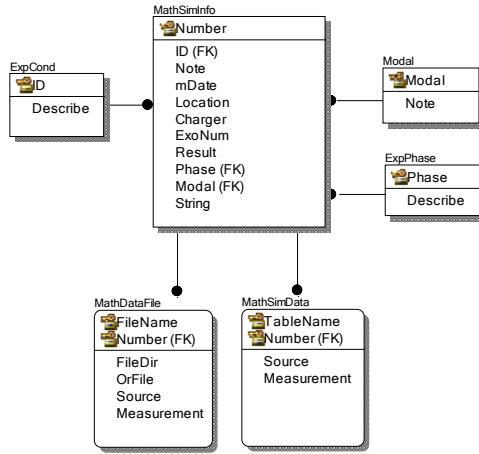


Fig. 2. The logical model of simulation result data

### Physical Model

Physical model is usually used to describe the structure of special DBMS, which transforms the entities of logical model to table, and transforms relationships to referential integrity. The physical model is depicted in figure 3.

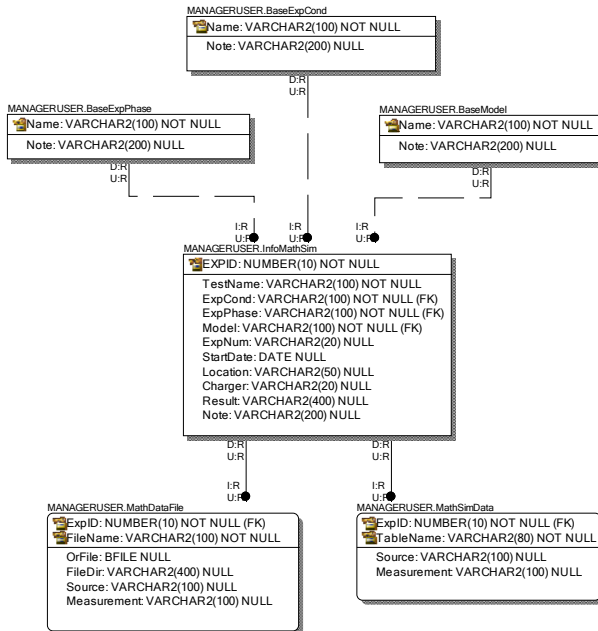


Fig. 3. The physical model of simulation result data

The referential integrity in physical model is depicted in table 1, which is efficient for preventing the exception of data insert, delete, and update operation, and also for improving the security of the data and the database.

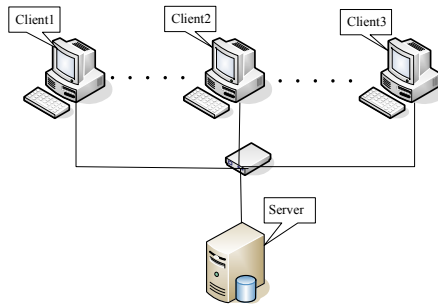
**Table 1.** The referential integrity in physical model

	<b>Insert</b>	<b>Delete</b>	<b>Update</b>
Parent Table	<b>None</b>	<b>Restrict</b>	<b>Cascade</b>
Child Table	<b>Restrict</b>	<b>None</b>	<b>Restrict</b>

In view of ERW tool, the corresponding SQL scripts would be generated based on the physical model mentioned above automatically. It would create tables and relationships between each other in Oracle10g by executing the SQL scripts.

### 4 Structure of Hardware

Resource management system will work on model of C/S. Its structure of hardware is depicted in figure 4.



**Fig. 4.** Structure of hardware for resource management system

### 5 Functions of Vrms

VRMS functional structure is depicted in figure 5.

It consists of two parts. One is underlying database; the other is resource management platform. Underlying database is used for storing all kinds of resource. Resource management platform provides the powerful tools to realize the function for user.

User&Role Management module: provides function for Authority distribution and Set careful and multi-level access to control strategy.

Test Information Management module: Management mathematical simulation and test, the hardware-in-the-loop simulation and test, the flight test and the data from the test equipment.

Basic Information Management module: Management model, test conditions, the missile test phase, the state and other basic information.

Resource Import/Export module: To provide convenient and rapid resources into output function, the greatest degree of improving resources into automation level.

Resource Query module: To provide the resources single, composite, the fuzzy inquiry, summary display.

Resource Backup module: Backup database table, view, and index of the object backup and restore function, improve the database security.

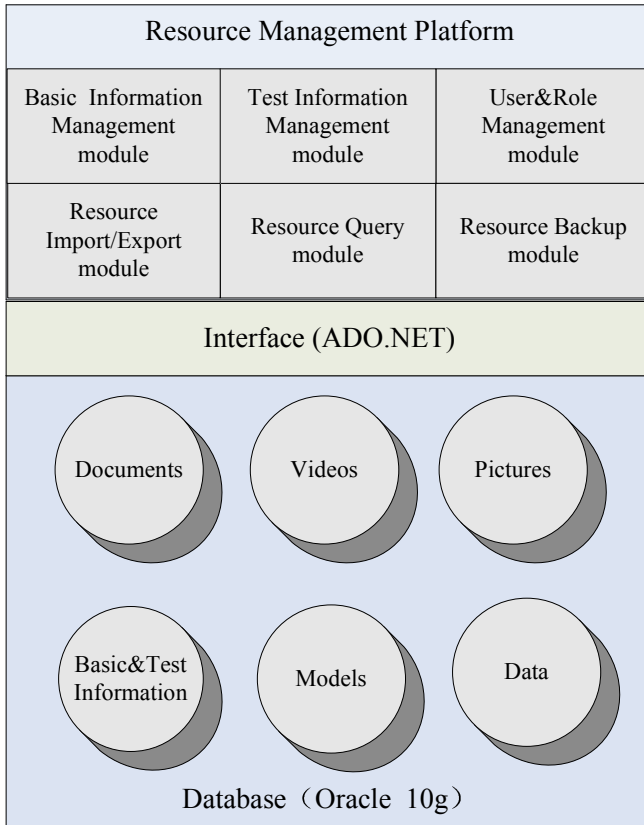


Fig. 5. Resource management system software structure

## 6 System Realization

In VS 2008 environment development the client application, the ado.net development between the client and the database of the interface between program.

The system finally used in local network. In order to improve the efficiency and safety of data exchange, the management system uses C/S mode.

Management system provides friendly, convenient and interactive platform for related workers and the underlying database interactive.

## 7 Summary

This paper established the guidance simulation system of VV&A resources logical model and physical model.

It is the one with strong reusability and a general design of reference to all sorts of guidance VV&A simulation system of the resource pool.

On the basis of the development of resources management system function is nearly perfect.

It realizes all kinds of resources and improves the management of resources for all the safety and VV&A automation level.

## References

1. DMSO Department of Defense Modeling and Simulation (M&S) Master Plan (DOD 500059-9P October 1995) 1995
2. Sargent, R.G.: Verification and Validation of Simulation Models. In: Proceedings of the 2007 Winter Simulation Conference, pp. 124–137 (2007)
3. Skoutas, D., Simitsis, A.: Ontology-based Conceptual Design of ETL Processes for both Structured and Semi-structured Data. *International Journal on Semantic Web and Information System (IJSWIS)* 4(3), 1–24 (2007)
4. Li, H.-Y., Chen, L.-F., et al.: System Test Database and Its Application in Simulation System. *Computer Simulation* 23(5), 5–9 (2006)
5. Liu, X., Shang, C.-Y.: Database Design and Application for Air Combat Simulation System. *Aeronautical Computing Technique* 36(2), 53–55 (2007)
6. Jiao, P., Tang, J.-B.: Result Validation Tool for Guidance Simulation. In: The International Conference on Computational Intelligence and Software Engineering, pp. 12–14 (2009)
7. MSCO. Documentation of Verification, Validation, and Accreditation (VV&A) For Models and Simulations (MIL-STD-3022) (EB/OL) (September 2008)
8. DMSO. Instruction 5000.61. DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (DoD Instruction 5000.61) (EB/OL) (2007), <http://www.dtic.mil/whs/directives/>
9. Cowdale, A.: Lessons Identified From Data Collection For Model Validation. In: Proceedings of the 2006 Winter Simulation Conference, pp. 1280–1285 (2006)



# Web-Based Collaborative Construction of Curriculum Texts in Higher Education

Wenxin Deng<sup>1</sup> and Fang Lu<sup>2,3</sup>

<sup>1</sup> Department of Informational Technology, Guangdong Teachers  
College of Foreign Language and Arts, Guangzhou, China

<sup>2</sup> College of Educational Information Technology,  
South China Normal University, Guangzhou, China

<sup>3</sup> Center of Educational Technology, South China University of Technology,  
Guangzhou, China

dengwx@gtcfla.cn, flu@scut.edu.cn

**Abstract.** This paper mainly discuss the relationship between four kinds of texts which proposed by Professor Zhong Qiquan and five curriculum forms which proposed by J. I. Goodlad, and propose four kind of method to the collaborative development of curriculum texts based on the case study. The collaborative construction mode include ready-made texts based on Wiki, teaching plan texts based on network teaching system, classroom texts based on online mind map website, and after-class texts based on blog.

Practice shows that, collaborative construction modes based on network can help complete curriculum texts construction in higher learning institution effectively, support curriculum teaching and learning activities, and improve teaching effect.

**Keywords:** Curriculum Text, Collaborative Construction, Network, higher learning institutions.

## 1 Introduction

### 1.1 The Curriculum Form Theory of J. I. Goodlad

Curriculum is the key of the development, research and reform for higher learning institutions in the field of education.

J. I. Goodlad, the famous scholar on curriculum in America, proposed that curriculum have five different forms in the light of curriculum forms, including ideological curriculum, formal curriculum, perceived curriculum, operational curriculum and experiential curriculum. [1]

### 1.2 The Typical Forms of Curriculum Text of Higher Learning Institutions

In the life cycle of the planning, design, implementation, evaluation and even revision and improvement of curriculums, there are various forms of expression, such as texts, graphics, symbols, etc, which are produced by the national and local education

administrations, schools, teachers and students. These forms of expression are mainly reflected in text forms, generally called curriculum texts. Curriculum texts are the crystallization of the practice of curriculum, the implementation of curriculum theory [2], and the extension of the curriculum contents. In this paper, curriculum texts refer to the various forms of texts.

Professor Zhong Qiquan, a famous expert in curriculum and teaching theory in the East China Normal University, proposed that curriculums could be simply texts, and four kinds of texts would be generated when the curriculums were implemented: ready-made texts, teaching plan texts, classroom texts and after-class texts.

### 1.3 The Corresponding Relationship between Curriculum Patterns and Curriculum Texts

Is there any corresponding relationship between curriculum patterns and curriculum texts? If the answer is yes, then what kind of relationship it will be?

**Table 1.** The corresponding relationship between curriculum patterns and curriculum texts

<b>Curriculum States</b>	<b>Explanation</b>	<b>Curriculum text</b>	<b>Express form of curriculum text</b>
ideological curriculum	It is setting by the promotion of curriculum research institutions, academic groups, curriculum experts	ready-made text	Initial curriculum standards, curriculum outline, etc
formal curriculum	It is specified by education management departments and schools		More detailed curriculum outline, teaching materials, teaching calendar, etc
perceived curriculum	It is the comprehension of teachers and students.	teaching plan text	Teaching design plans, lesson plans, PPT electronic presentations, materials such as pictures, videos and so on, structured resources like courseware and so on

**Table 1.** (continued)

operational curriculum	It is implemented just in the classroom	classroom text	Texts which formed during the implementation of curriculum, such as Language, blackboard-writing, classroom activities, video of class, etc
experiential curriculum	Experience of teachers and students after class	After-class text	Teaching narrative, research papers, blog, lecture notes, homework, etc.

Through the observation of the linear curriculum formation process described by J. I. Goodlad and the different forms of typical curriculum texts of higher learning institutions, we have found that there exists a corresponding relationship between curriculum patterns and curriculum texts, as shown in Tab.1.

Among them, the ideal curriculum and formal curriculum correspond to the "ready-made text"; the comprehended curriculum corresponds to the "teaching plan text"; the implemented curriculum corresponds to the "classroom text," and the experienced curriculum corresponds to the "after-class text" [4].

## 2 Methods to Develop Curriculum Texts

The process of curriculum from its target setting to implementation is the process of changes of curriculum texts forms. The construction of curriculum texts plays a key role in the construction of curriculum.

### 2.1 The Main Methods of Curriculum Texts Development

Curriculum texts development involves planning, organization, implementation, evaluation and revision of teaching content and teaching activities in a certain field which chosen according to curriculum objectives and defined after demand analysis. There are mainly two methods of curriculum texts development: lifecycle method and rapid prototyping method.

(1) Lifecycle method. It also called structured system development method. It is a popular information system development method both at home and abroad now, widely used and promoted in system development. Especially when it is used to develop complicated large-scale systems, this method displays incomparable superiority. Using this method for reference in curriculum texts development, there are mainly four phases: demand analysis, design, implementation and operation.

(2) Rapid prototyping method. Lifecycle method's outstanding advantage is its emphasis on integrity and wholeness through the system development process, ensuring the quality of curriculum texts development. However, lifecycle method takes a long development period. In order to shorten the development cycle, rapid prototyping method was introduced. Rapid prototyping method was first used in mould manufacturing. Afterwards it is widely used in system development, including curriculum construction. During the period of curriculum texts developments, four aspects are emphasized in rapid prototyping method: objective orientation, user participation, uses of tools and constant perfection.

## **2.2 Basic Methods of Curriculum Texts Development**

The basic methods of curriculum texts development include independent development and cooperation development, the latter is the mainstream way currently.

The important guidance documents such as Curriculum standards, textbooks, teaching design program, will be difficult to accomplish without cooperative. Even some texts which seem very personalization, such as texts writing on the blackboard, teaching narrative and teaching reflection, also can be cooperative construct.

## **2.3 The Main Problems of Curriculum Texts Development**

- (1) Long development cycle;
- (2) Difficult to control the development process;
- (3) Stereotyped forms of expression of the curriculum text;
- (4) Difficult to share.

# **3 Web-Based Collaborative Construction Modes of Curriculum Texts**

The rapid development of information technology comprised mainly of computer and network technology, and the in-depth implementation of network technology and Web2.0 in the field of higher education, has led to a reform in the form, content, and method in curriculum texts construction and provide more strong supports for the development of curriculum texts. The collaborative construction and sharing of curriculum texts are the keys in the reforms in curriculum construction.

The collaborative construction of web-based curriculum texts in the higher learning institution refers to the collaboration between teachers, between students, and between teacher and students and a sharing of learning resources. There are four main collaborative construction modes as outlined below.

## **3.1 Wiki-based Collaborative Construction Mode of Ready-Made Texts**

Ready-made texts are forms of texts of ideological curriculum and formal curriculum, and they are the base for teacher to comprehend and reconstruct curriculum. When an ideological curriculum is to be established as a formal curriculum, normally the syllabus will be firstly finalized top-down (planned by the administration departments) or bottom-up (planned by teachers and approved by the administration

departments), and then teacher will choose or compile the textbook based on that syllabus.

The textbook, as the most important form of text, is a manifestation of teacher's latent knowledge, the important fruit of curriculum construction and resources for teaching activities.

Since the concept of Web2.0 was created in 2004, Wiki technology as one of the important application of Web2.0 has been widely used in collaborative writing. For example, Wikipedia (<http://www.wikipedia.org/>) is a collaborative encyclopedic website which is free, of no charge, and with open contents based on Wiki technology. Baidu encyclopedia (<http://baike.baidu.com/>) is also a well-known encyclopedic website in China. It makes possible to define and interpret various words and terms with collaborative efforts in edition.

In 2009, we adopted Wiki technology in a collaborative compilation of a textbook *Introduction to Modern distance education learning*. The outcome shows that with the help of Wiki, the quality and efficiency in compilation could be significantly improved.

### 3.1.1 The Construction of Collaborative Writing System Based on Wiki Technology

(1) First, we should choose an appropriate server and use PHP + MySQL + Apache technology solutions to install the operating system and software system of the server. There are some enterprises provide server space free of charge or rented it to the user. The compilation of teaching textbooks usually has timeliness, less demand for capacity, few users and concurrent users. Thus, this approach was adopted in the absence of sufficient technical strength.

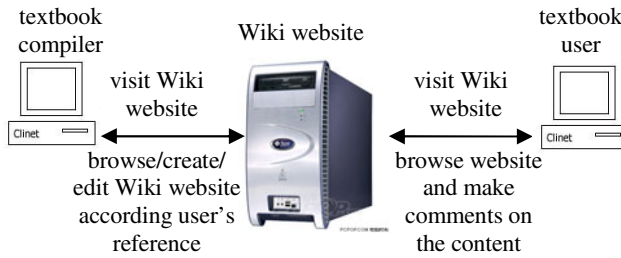
(2) Download the free Media Wiki installation program from the Wiki website, and then install the program and configure the server according to the instruction of the program.

(3) Enter the URL <http://localhost/index.php> (local host is the server host's IP address) in the browser address bar, you can visit the Wiki website that has been built. If necessary, you can apply for domain names.

3.1.2 The textbook compilation team built the framework of textbook and compile textbook on the Wiki website according to their assignment.

3.1.3 Every teacher (team member) can read the contents of the others, make annotation, revision and so on, and each move will be recorded by Wiki website as the reference for all the team members. The chief compilers can accept or put forward their own views and discuss with the one that make the comment.

3.1.4 Invite the users of textbook (such as students) to read the textbook and evaluate it from the user's point of view, such as to see the content of the textbook is clearly elucidated.



**Fig. 1.** Collaborative compilation system based on Wiki technology

In the collaboration between compilers (Teacher), peers (other teachers), user (learner), the compilation of textbook can be improved from different perspectives, and so its completion with high quality and efficiency can be ensured.

### 3.2 Collaborative Construction mode of Teaching Plan Texts Based on Network Teaching System

Teaching plan text is the text form of comprehended curriculum, and used in the teaching and learning activities for teacher and student. The process of teaching plan text's formation is accompanied by the comprehension of curriculum by the teachers and students.

#### (1) The main type of teaching plan texts

Teaching plan text includes lesson plans, case study, teaching materials, courseware, online courses and special learning websites. PPT presentation document is the most frequently used and the most important text in the form of teaching plan text. In addition, with the range of electronic resources and multimedia content growing wider and wider, teaching plan texts also include graphic images, animation (2D, 3D, virtual reality, etc.), audio and video (single-screen video, three-screen video, etc.) materials, and other electronic resources such as courseware (CD-ROM, networking courseware); web-based courses; online courses. All these types of multimedia resources complement each other in some extent.

#### (2) The collaborative construction of teaching plan text based on network teaching system

With the development of E-Learning, teaching plan text gradually extends from an independent resource to the comprehensive resource database, course website (the most typical course website provides quality courses, as shown in Fig 2) and on-line courses based on network teaching system (Fig 3) which integrate all types of teaching plan text with a uniform management and application platform. Teachers, teaching assistants, peers, and students are granted with different levels of access rights to the platform. They build, evaluate and improve the teaching plan text collaboratively on the course platform.



Fig. 2. The website of national quality course *The Foundation of Machine Design*



Fig. 3. Online course *High-level Language Programming* based on online learning platform

### 3.3 Collaborative Construction of Classroom Text Based on Online Mind Map Website

Classroom text is the text form of curriculum in operation, and the writing on the blackboard is one of its typical forms. Through the use of electronic whiteboards, pens and other writing tools for computer input, handwriting functions of the PPT software, the written text can be saved into the notes page of the PPT document, or stored on the computer in the form of text file. Classroom text is the sparks in the teaching process and in the communication between teachers and students in class, and it is a good supplement to curriculum texts.

Mind Map is an excellent tool for knowledge representation. It can effectively externalize tacit knowledge. Mind Map can be draw mainly through hand-painting, mind map making tools, online mind map site.

Text in the classroom generate by students is equally important with the text generate by teachers. Teachers and students through the online mind map website (like <http://www.imaginationcubed.com/>) cooperate with each other online on certain research topic, draw mind map together in the form of brainstorming, and output the classroom text in images or web pages. This learning method has a good effect on the development of the spirit of innovation and collaborative abilities of students.



**Fig. 4.** Teachers and students draw mind map with online mind map website

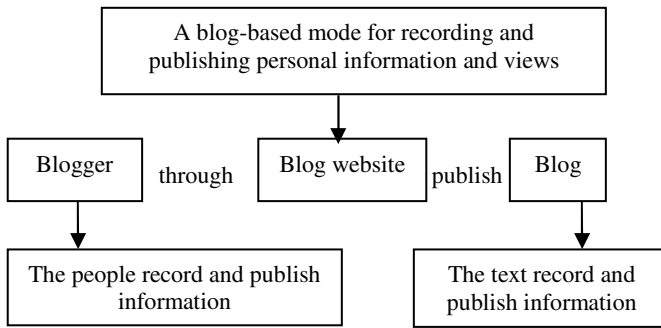
### 3.4 Collaborative Construction of After-Class Text Based on Blog

After-class text is the form of implemented curriculum. Deep thought, can quick. Teachers' reflection is very important for improvement of the curriculum. Only when teachers continually reflect on the quality of the implemented curriculum can they provide cases, inspiration and hints for the reconstruction of the formal curriculum, and improve the teacher's comprehension of the curriculum [5]. Teaching notes, teaching cases, and teaching narratives are the forms of after-class text. They are the narration of the thinking, behavior, and emotion of the teacher and student in teaching practice with little difference among the three. They all belong to qualitative research method where the researcher himself is a research tool, and adopt the method of induction to analyze data and form theory in a natural context.

Teaching blog (Blog) is an important development tool of after-class text. Now, there are a number of educators, educational institutions create Blog, such as the blog "Eastbound Records" created by the famous educational technology expert Li Jiahou, and the blog "Tianhe Tribes" by the Department of Education for primary and secondary schools in Tianhe District in Guangzhou. These blogs have built up teaching and research community, or learning community in different regions and groups of people, and significantly promoted the reflection on and reform of teaching.

Blog provides a convenient way to release personal information of the network. Teacher should first select a blog website, such as Sina blog, Netease blog, Blog Bus, etc. and register in the blog website, and then choose website template or design it by yourself, and finally publish his own teaching materials on the blog.





**Fig. 5.** The working process of blog

Lin Chongde, a famous education expert, pointed out that the development formula of outstanding teachers is “good teachers equals teaching process plus reflection”. [6] Reflection raises questions on the implemented curriculum, and it is an effective way to create the learning organization. Currently, many Chinese universities and colleges have developed relevant mechanisms to encourage teachers to create blogs, published papers and other forms of narrative teaching, teaching cases and other fruits of teaching reflection, so as to promote team learning and professionalism of teacher.

New forms of technology prompt the reform of the media, which bring forth more new types of curriculum text. For example, texts generated at the Online Q&A and network forum are a new type of both implemented curriculum texts and experienced curriculum texts. Now, using curriculum website and online courses to create, manage and apply curriculum texts is the important method of systematic construction of curriculum texts.

## 4 Conclusion

Five forms of curriculum which proposed by Goodlad and the corresponding four curriculum texts constitute a unique perspective to evaluate the construction of university curriculum. The collaborative construction of curriculum texts is the most important part in the development of curriculum. From it, we can see the essence and uniqueness properties of curriculum text. The expression forms of Curriculum is text, that reveals a sense of self-recognition. [7] Teachers and students should understand the course text in the process of understanding life, find its meaning, expand their horizons, and gain useful knowledge from the alien world during the process of perceiving curriculum texts. [8] We should not turn curriculum into the text mechanized in the development of Curriculum. We should perceive the curriculum and express in text form. Texts were generated dynamically and continually, and finally reconstruct curriculum in forms of text. The subsequent teaching experiments showed that these web-based curriculum texts collaborative construction mode can help the construction of curriculum effectively, support learning and teaching activities, and prompt teaching effect.

We believe that the rapid development of network technology will provide more collaborative and more effective support for the collaborative construction of curriculum text.

## References

1. Goodlad, J.I.: Curriculum Inquiry: The Study of Curriculum Practice, pp. 60–64. Mc Graw-Hill book company, New York (1979)
2. Shi, T.: Text Conversion in Curriculum implementation. Education Tribune (1), 15–17 (2003)
3. Zhong, Q.: Dialogue and Texts: The transformation of the teaching standard. Educational Research (3), 33–39 (2001)
4. Du, Z.: Research of the Comprehend course, p. 84. Guangming Daily Press (2003)
5. Shi, T.: Text Conversion in Curriculum implementation. Education Tribune (1), 15–17 (2003)
6. Xin, T., Shen, J.: Research of Teacher's Instructional Monitoring capability. Journal of Beijing Normal University (Social Science Edition) (1), 67–75 (1996)
7. Wang, Y.: Read the course text in comprehension paradigm. Exploration of Education (9), 33–35 (2002)
8. Gao, X.: Understanding-based teaching: A new concept of environmental education. Green Education Lecture Notes (1), 9–10 (2009)

# Study on LDPC Application in Wireless Communication System

Li Wang, Tikui Wang, Jianqiang Guo, Jinlong Li, and Xiaorong Gao

Photoelectric Engineering Institute, Southwest Jiaotong University, Chengdu, 610031  
553699304@qq.com

**Abstract.** Low-density-parity-check (LDPC) is linear group code with sparse check-up matrix. It not only has the good performance of approaching the Shannon limit, but also lowers decoding complexity and the structure is more flexible. This paper mainly focuses on LDPC whose length is 672 bit and rate is 2/3 in the IEEE802.16e standard, and completes LDPC simulation in wireless communication system using Matlab and Simulink. The conclusion shows that LDPC has good performance in a simulation atmosphere noise condition.

**Keywords:** LDPC, group code, simulink, wireless communication.

## 1 Introduction

The purpose of the communication is timely and reliable transmitting the information to each other. The concern of communication system in design is how to make the information transmission error rate the lowest when data source power and bandwidth are limited, and the cost of system complexity and equipment should be as small as possible.

LDPC was put forward by Gallager in 1962. By iteration decoding, the bit error probability decreases in an exponential way when the code length is increasing [1]. With recent years researching of LDPC, LDPC has a very good resistance to decline and higher encoding gain. This code has more advantages than Turbo code, and can fulfill the future communication system with high-speed data transmission and performance. It is most probable to become application solutions of the fourth generation mobile communication system [2].

## 2 Algorithm

LDPC encoding pattern has direct code and fast code. The direct code is similar to general linear group code encoding algorithm, it needs to calculate generator matrix  $G$ . We can get coding sequence through  $c = m \times G$ . But the multiplication complexity of this matrix is  $O(n^2)$  which is linearly proportional to the square of code length. As a result, the storage for generating matrix consumes a lot of resources that if code length is longer, computational complexity will be huge. Realizing this is of great difficulty.

Fast encoding algorithm is direct to information bits ( $s$ ) encoding by pretreatment block matrix H. Fig 1 shows matrix H pretreatment block.

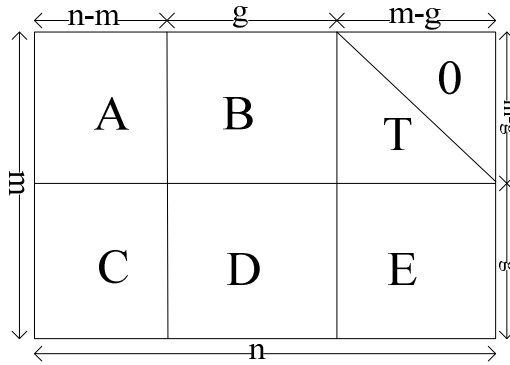


Fig. 1. Check-up matrix block

According to the following formula,  $p_1, p_2$  can be calculated as

$$p_1^T = -\partial^{-1}(-ET^{-1}A + C)s^T \tag{1}$$

$$p_2^T = -T^{-1}(As^T + Bp_1^T) \tag{2}$$

$$Hc^T = \begin{bmatrix} A & B & T \\ C & D & E \end{bmatrix} [s^T \ p_1^T \ p_2^T]^T = 0^T \tag{3}$$

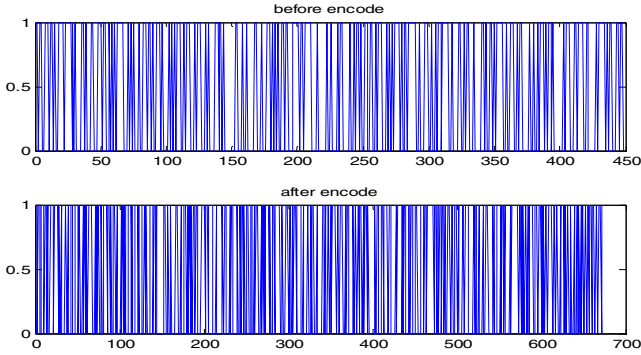
$$\begin{bmatrix} A & B & T \\ C & D & E \end{bmatrix} \begin{bmatrix} s^T \\ p_1^T \\ p_2^T \end{bmatrix} = 0^T \tag{4}$$

$$As^T + Bp_1^T + Tp_2^T = 0 \tag{5}$$

$$Cs^T + Dp_1^T + Ep_2^T = 0 \tag{6}$$

Where define  $\partial = -ET^{-1}B + D$ . Encode code  $c$  can be got through  $C = (s, p_1, p_2)$ .

In Matlab, this paper finishes fast encoding algorithm simulation for LDPC of code length 672 bit, line weight 9 and column weight 3.



**Fig. 2.** Coding wave figures before and after

Comparing the two figures above, we can see the original data code length is 448 bits while that becomes 672bits after encoded. Encoding efficiency is  $2/3$ . The results are conforming to the simulation parameter settings.

### 3 Wireless Channel Simulink Simulation

After LDPC encoding, the output encoding data will be the signal source of system simulation. BPSK is one of the most common modulation modes. Supposing baseband code is  $m(t)$  and carrier frequency is  $w_0$ , then phase-modulation wave can be expressed as

$$s(t) = \cos[w_0 t + \partial_{c(t)}] .$$

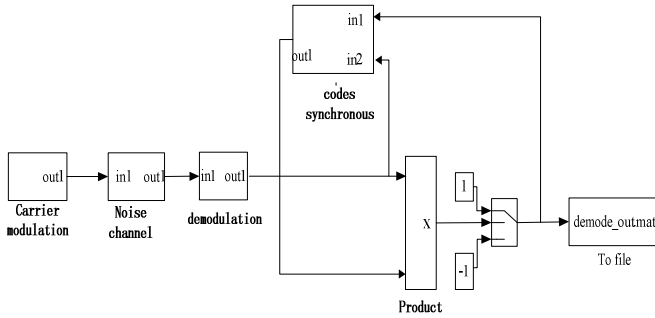
Where define  $\partial$  is phase modulation coefficient. In the practical application, baseband usually use dual polarity codes, that is  $m(t) = \{-1, +1\}$ , so BPSK modulating signal can be expressed as

$$s(t) = m(t) \cos[w_0 t + \partial]$$

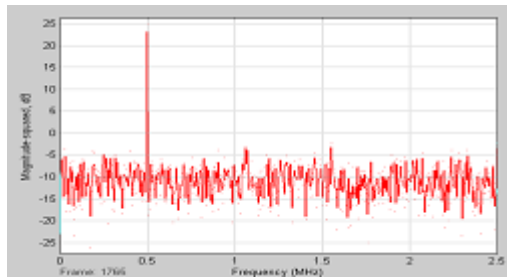
In this paper, AWGN and pulse noise [4] [5] (similar to the atmospheric noise) are concerned, which are simulated in the channel. The simulation system scheme is shown in fig. 3.

From Fig.4, it can be seen that spectrum distribution is quite uniform and regular. It shows that the AWGN and pulse noise already mix to the channel.

In digital communication system, synchronization is a key problem, which includes carrier synchronization, code element synchronization, group synchronization and nets synchronization. The paper focuses on code element synchronization.

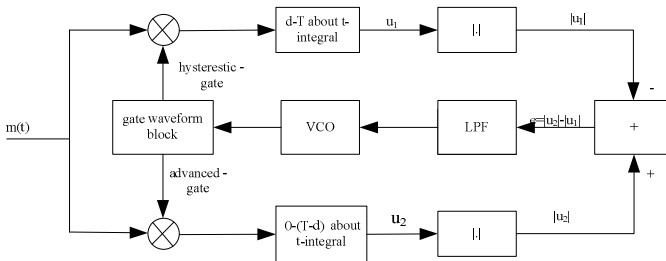


**Fig. 3.** Channel simulation subsystem model

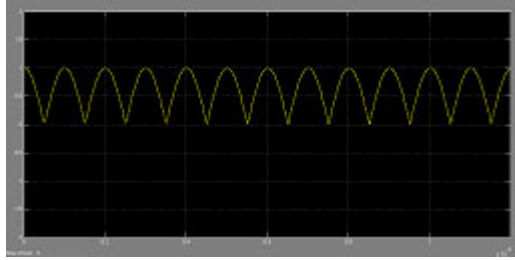


**Fig. 4.** Noise spectrum

When receiving digital signals, in order to judge the receiving codes element and energy integral of receiving yards element accurately in correct moment, we must know accurate begin-end moment of receiving codes element. Receiving codes element begin-end moment information can be got from synchronization pulse sequence [6]. One of the most common method is advanced/lag up door constituent code element synchronization.



**Fig. 5.** Advanced/lag up door synchronization schematic diagram



**Fig. 6.** Code element synchronous sampling clock

From the graph after reforming, clock signal extracted accords with the sampling and vindicating clock.

## 4 LDPC Decode Design

LDPC decoder adopts iterative decoding algorithm, and the most basic is BP algorithm. It is a kind of soft decision iterative decoding algorithm based on probability. BP algorithm includes five steps: 1. Initialization; 2. Variable node update; 3. Calibration node update; 4. Decode judgments; 5. Calibration. [7]

LDPC decoder module design and simulation.

### 1. Simulation

- (1) Load signal file (demode\_out.mat) after codes element synchronous sampling decision;
- (2) Make use of program in main\_decode.m to load the check-up matrix information and invoke program in decode.m.

### 2. Initial parameters of the simulation

The simulation uses regular LDPC (672, 3, 9), and the signal-to-noise ratio of channel is 10dB. In the simulation, we need to invoke two programs (decode.m and main\_decode1.m).

```
>> main_decode1

derr =

     1

err =

     0
```

**Fig. 7.** Decoding the output

From the graph above, we can see a mistake appears after signal encoded passes through noise channel. But finally the mistake is corrected by LDPC decoding algorithm, which proves that LDPC is error correcting code.

## 5 Conclusion

The paper studies on LDPC's error-code ratio performance. Fast encoder algorithm, code element synchronization, and the BP decoding algorithm are used in the paper. Within Matlab and Simulink, the performance of the LDPC is analyzed systematically, which confirms that LDPC performs very well under the influence of simulate atmosphere noise. Because of LDPC's superior function, it could be applied in wireless optical communication field. With the development of FPGA, the hardware realizing longer LDPC will be feasible.

**Acknowledgments.** Wang Tikui, 1987, master, researching direction is wireless communication.

## References

1. Gallager, R.G.: Low-density parity-check codes. IRE Trans. on Inform. Theory 8(1), 21–28 (1962)
2. Mackay, D.J., Neal, R.M.: Shannon limit performance of low density parity check codes. Electronics Letters 3(18), 1645–1646 (1996)
3. Li, X.: A kind of applicable to the LDPC code design based on FPGA. Aviation Electronic Technology 3(2), 15–19 (2006)
4. Zhang, W., Wang, Y.: Vlf-em/low frequency noise simulation research of communication atmosphere. Ship Electronic Engineering 26(6), 124–125 (2006)
5. Li, X., Zhu, L.: Atmospheric conditions/LF noise compared communication system simulation. Space Electronic Technology 3, 7–8 (2010)
6. Fan, C., Cao, L.: Communication principle, vol. 3(6), pp. 413–414. China National Defense Industry Press, Beijing (2008)
7. Zhao, G.: Spread spectrum communication system is practical simulation technology, vol. 10, pp. 164–165. China National Defence Industry Press (2009)



# Coding of Magnetic Ruler in Cylinder

Gengbiao Chen and Yang Yang

Advanced Vocational Technical College, Shanghai University of Engineering Science,  
Shanghai 200437, China

cupiter@126.com, yydaisy@163.com

**Abstract.** One of magnetic rod hydraulic cylinder displacement measurement system's major techniques is to carry on "the rod" according to certain coding rule scoring, the good code not only need be advantageous to the recognition time the error detection or the error correction, moreover must be advantageous to the long traveling schedule survey time the symbol expansion, the immediate influence to the hydraulic cylinder piston displacement's effective survey, this article has analyzed the magnetic rod code design fit and unfit quality weight target - - The average amount of information sources, the decoding error rate, error detection/error correction ability, has carried on the comparison based on the target to the existing magnetic rod code, had pointed out their good and bad points, provide the basis for the magnetic rod's efficient coding.

**Keywords:** cylinder, magnetic ruler, coding.

## 1 Introduction

The working mechanism of the magnetic ruler in cylinder refers that the flat-end groove is produced on the cylinder piston rod which is the steel material and that has been specially processed according to "certain rules". Then the ceramic coating is covered. When the cylinder runs, the magnetic field will be established around the flat-end groove. And it is required to employ the magnetic sensitive components to measure the magnetic changes around the flat-end groove so as to obtain the stroke value. The magnetic ruler measurement [1][2] has the following advantages: there is no component wear problems due to the relative motion; the information sensitivity is high and the dynamic response is good; the integration and the intelligence of sensors are easy to be implemented; the power consumption is low and it is safe and reliable, etc.

The "certain rules" on magnetic ruler essentially refer to the coding rules. The good coding is not only conducive to the error detection or the error correction in the process of identification, but also helpful for the code word expansion in the long stroke measurement. This has directly affected the effective measurement of the cylinder piston displacement. In this paper, the indicators of the pros and cons of magnetic ruler coding design have been analyzed at length from the perspective of non-contact displacement measurement. At the same time, the advantages and the disadvantages of the existing several magnetic ruler coding have been compared and the design as well as the optimization methods of good coding has been also briefly stated.

## 2 Analysis of the Indicators of Magnetic Ruler Coding

From the principles of the displacement measurement of cylinder magnetic ruler, we can find that the information is transmitted according to the width changes and the vertical arrangement of “boss” and “groove” of the magnetic ruler which corresponds to the code “1” and “0”. The coding rules are directly related to the measurement accuracy. Therefore, it is very necessary to study the evaluation indicators of the pros and cons of coding design.

From literature [3], we can find that the information density of bar code, the resolution and the error correction ability are the main indicators to evaluate the pros and cons of the black and white bar code design under the photoelectrical detection, which has provided a certain reference for the measurement of magnetic ruler coding under the magnetic detection. However, due to the different use media and the detection methods of the two codes, the indicators of the pros and cons of magnetic ruler coding have the different characteristics, which have been analyzed as follows.

### 2.1 The Average Amount of Information in Source

Assuming that the discrete information source is a collection of  $n$  symbols and each symbol  $x_i$  independently occurs in information according to the probability  $P(x_i)$ . Then the average amount of information is

$$H(x) = -\sum_{i=1}^n P(x_i) \log_2(P(x_i)) \quad (\text{bit/symbol}) \quad (1)$$

### 2.2 The Error Probability of Coding

As for the coding which employs the relevant encoding algorithm, an error will affect a sequence or the fragments of sequence. If the sending sequence is inconsistent with the receiving sequence, the wrong time will be presented. As the sequence has a certain length and it maybe have a variety of values, so it is impossible to analyze every mistake. If we employ the linear code, then the problems will be simplified according to the linearity.

Assuming that the sent sequences are all-zero sequences, then the correct coding sequence should be the all-zero sequence path on the top of lattice diagram and this path is a correct one. Any coding sequence which deviates from this correct path is the wrong path. The process that the sequence deviated from the correct path at certain moment and then came back after several steps is considered as the error event. The corresponding path is the error path. Based on the concept of free distance, the distance between any error path and correct path must not be less than the free distance and the specific value is uncertain. The error probability under the condition of the binary symmetric channel and the hard decision has been analyzed in the following part.

In the Viterbi coding algorithm, if the distance between a certain error path and the receiving sequence is less than the one between the correct path and the receiving sequence, the decoder will select the error path as the maximum likelihood path. Then the coding will be wrong. Assuming that the weight of error path is  $d$ , when the

coding is wrong, then the weight of receiving sequence must be greater than  $d/2$ . Therefore, the condition of error event refers that the weight of receiving sequence must be within  $[(d+1)/2, d]$ , then the error probability is

$$p(E, d) = \sum_{l=(d+1)/2}^d \binom{d}{l} p^l (1-p)^{d-l} \quad (2)$$

Considering  $p$  is small, when  $l \geq (d+1)/2$ , the above formula can be written as

$$p(E, d) \approx \left( \sqrt{4p(1-p)} \right)^d \quad (3)$$

### 2.3 The Ability of Error Detection/Error Correction

In the code-block collection of coding, the minimum distance between any two available code groups is called the minimum code distance  $d_{\min}$ .  $d_{\min}$  is greater, which will indicate that the difference between the two code words is greater and the error detection ability as well as the error correction ability is stronger. The minimum code distance is an important parameter of coding, which is directly related to the error detection ability and the error correction ability.

As for the block codes, the minimum distance should meet the following formula.

$$d_{\min} \geq \begin{cases} e+1 & \text{When } e \text{ errors are detected} \\ 2t+1 & \text{When } t \text{ errors are corrected} \\ t+e+1 & \text{When } t \text{ errors are corrected and } e \text{ errors} \\ & \text{are detected} \end{cases}$$

## 3 The Comparison of the Existing Codes

The comparison of the most commonly used magnetic ruler coding methods is analyzed as follows.

### 3.1 The Coding That Imitates the Serial Communication

The coding that imitates the serial communication refers to the phenomenon that the binary sequence of magnetic ruler will be composed as the frames according to the asynchronous communication, which takes the serial asynchronous transfer mode of computers as the reference. Then each frame is assigned to different data and the absolute coding value will be read through the data bits. The key of this coding method is to set the start bit with appropriate bits in a complete frame. The start bits are more, and it will be easier to distinguish it from other bits. But the length of each frame will become larger and the read of coding numerical value will take much more asynchronous transmission time. Meanwhile, when we only use the start bit as the flag, it is required to remove the frames which maybe cause the confusion for eliminating the frame interference, which will greatly reduce the coding efficiency.

### 3.2 The Frame Superimposed Coding

The data transmission method of frame superimposed coding can solve the inconvenient operation and the low efficiency of the coding that imitates the serial communication method. It has directly regarded any bit in frames as the start bit. When the length of a frame is reached, the binary sequence will be the absolute coding numerical value [4]. This coding method does not employ the start bit, the parity bit and the stop bit. At the same time, there is only a change in adjacent frames and the other bits are arranged in overlapping. Each change will correspond to a coding value, which has avoided the interference of frames. However, as for this coding method, the reading and the measurement can be only implemented after a large distance. Meanwhile, its coding sequence is not unique, which is not conducive to the standardization expansion of the coding.

### 3.3 The Gray-R Coding

The Gray-R coding has employed the Gray as the numerical code. A fixed code width has been considered as the reference code R and the reference code is used to position the ruler and improve the accuracy. The Gray code is a non-weighted code with absolute coding method. The cycle and single-step features can eliminate the possibility of significant errors when the number is randomly obtained. In the transition between any two adjacent integers, there is only one bit changing, which has greatly reduced the logic confusion from one state to the next state and that has a strong fault tolerance ability. If the measurement range needs to be extended, we can employ the  $(n-1)^{\text{th}}$  Gray code area, the  $(n-2)^{\text{th}}$  Gray code area and some others after using n Gray code areas. Therefore, the measurement range can be doubled without increasing the sensors.

## 4 Conclusion

In this paper, the evaluation indicators of magnetic ruler coding have been analyzed and the existing three coding methods have been also compared. From it, we can find that the serial coding can not fully realize the absolute measurement and the coding efficiency has been greatly reduced due to the interference of frames. However, the Gray-R coding has employed the coding and decoding idea that imitates the parallel communication, which not only solves the problems of frame interference, but also greatly improves the anti-jamming ability of coding and decoding system.

**Acknowledgments.** This work is supported by the special research project for the selection and the training of outstanding young teachers in Shanghai universities (Grant No.: gjd09041).

## References

1. Chen, G., Wu, J., Wang, D.: Encoding and Recognition in Displacement Measurement Based on Magneto-dependent Sensor. *Journal of Electronic Measurement and Instrument* (22)(Suppl.), 483–485 (2008)
2. Chen, G., Zhu, S.: Gray-R Code Recognition of Magnetic ruler in Cylinder. In: *International Conference on wireless Networks and Information Systems, WNIS 2010*, vol. 2, pp. 175–177 (December 2010)
3. Chen, G., Wu, J., Le, Y.: The Study of Encoding Principle with Digital Levels. *Acta Geodaetica Cartographica Sinica* 37(3), 380–383 (2008)
4. Pan, F., Ding, F.: Research on Frame Superimposed Scale Coding and the Application on Stroke Sensing Cylinder. *Chinese Journal of Scientific Instrument* 24(4), 364–367 (2003)

# Edge Detection of Level Ruler with Digital Level

Gengbiao Chen and Caiyi Wang

Advanced Vocational Technical College, Shanghai University of Engineering Science,  
Shanghai 200437, China

cupiter@126.com, windyi\_jackal@yahoo.com.cn

**Abstract.** The digital level displacement surveys, the rod barcode image's marginal check are affects its measuring accuracy the important link. According to the rod barcode image's rectangular characteristic, proposed that uses based on the LOG operator zero crossing intersection examination algorithm, analyzed in the examination process to have "the false peripheral point" the question, gave three kinds to amend a plan, a plan use peripheral point's first derivative was the extreme value characteristic, the plan two use peripheral point three step derivative non-vanishing characteristic, about the plan three use peripheral points the domain product non-characteristic. The experimental verification three methods have gone to "the false peripheral point" validity.

**Keywords:** Digital level, edge detect, LOG operator, zero crossing.

## 1 Introduction

The measurement principle of digital level refers that the barcode pattern in rulers is reflected by light and a part of light beam is directly imaged on the partition board of telescope for visual aiming and focusing; however, another part of light beam is transmitted to the CCD sensor through the beam splitter, which has been converted into the digital signals by photo electricity and A/D. They are decoded by the microprocessor DSP. At the same time, they have also compared with the reference signals of the memory of instruments so as to get the high value of the barcode pattern in CCD[1]. The automatic reading of the traditional scale needs the array image sensor and it is required to identify the symbols of numeral. So the algorithm is complex and difficult. We just need the one-dimensional image sensor to replace the traditional ruler with the barcode ruler, and then the structure as well as the algorithm will be greatly simplified. In the edge detection algorithm of one-dimensional image[2], the zero intersection algorithm based on the LOG operator has the good detection effect. But it has the defect of "pseudo edge point". In this paper, three different revision methods have been proposed through studying the causes.

## 2 Log Operator

The LOG edge detection operator refers to the operator which smoothes at first and then computes the derivatives. It has firstly smoothed the images by employing the

Gaussian function and then it is necessary to carry out the Tabatabai computing for the filtered images[3][4]. Most of the classic operators are the first-order differential ones which are related to the directions. And the Tabatabai operator is a second-order differential operator, which simultaneously has the partial maximum of the first derivative and the zero intersection of the second derivative. Therefore, it has nothing to do with the direction of edge point. As for the dark area on the edge, the second derivative is positive. As for the light-colored areas, the second derivative is negative. And for the gray-continuous areas, the second derivative is zero. Thus, there is a zero intersection among the points with gradation changes, namely the edge point. As for the two-dimensional image signal  $f(x, y)$ , Marr has proposed that it is necessary to employ the Gaussian function to carry out the smoothing at first and then

$$g(x, y) = f(x, y) * G(x, y, \sigma) \quad (1)$$

In this formula,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\pi\sigma^2}\right) \quad (2)$$

Then it is required to strike the edge of the image function  $g(x, y)$  after the smoothing.

Corresponding to the processed discretization, it is necessary to employ the differential approach to take the sum of the second-order difference on the  $x$  axis and the  $y$  axis for each pixel  $(x, y)$  of digital image  $f(x, y)$ , i.e.

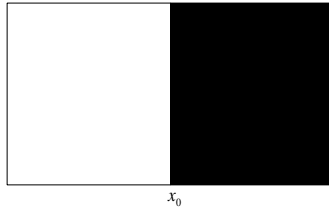
$$\nabla^2 F(X, Y) = f(x+1, y) + f(x-1, y) + f(x, y-1) + f(x, y+1) - 4f(x, y)$$

The LOG operator has the good edge detection effect. However, as the second derivative is more sensitive to the noise, so the anti-interference ability is not strong. In order to achieve the desired edge detection effect, the further improvement is required.

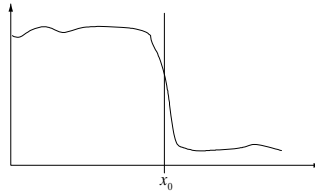
### 3 The Improvement of Log Operator

Due to the interference, the LOG operator has detected that the points whose second derivative is zero in the gradient images are not all the edge points and the causes are analyzed in the following part.

The Figure 1(a) refers to the digital image with step edges due to the grayscale mutation. And the results smoothed by the Gaussian function can be shown in Figure 1(b).



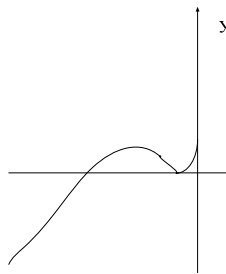
**Fig. 1.** (a) The barcode image before smoothing



**Fig. 1.** (b) The results after smoothing

Based on the zero intersection detection theory, it is required to compute the first derivative and the second derivative of the above function images. And a local extremum will be formed at  $x_0$ , i.e.  $f'(x_0) = \max f'(x)$ , in which  $x_0 \in (x_0 - \sigma, x_0 + \sigma)$  and  $\sigma \in R$ . The second derivative is the derivative of the local extremum, so  $f''(x_0) = 0$ .

If there is interference, the second derivative image of the image signal can be shown in Figure 2. And there is also a zero around the edge point due to the interference.



**Fig. 2.** The second derivative image with interference

The zero crossing point must cross the horizontal axis from the point which is greater than zero to the point which is less than zero, or vice versa. However, the first derivative of the “pseudo edge point” must not be the extremum. On account of this feature, three methods can be employed to remove them.



Method 1: the images whose second derivative is zero should be compared with the ones whose first derivative is the extremum so as to remove the points whose first derivative is not the extremum. The concrete steps are:

(1) It is necessary to employ the LOG filter to process the original images and the processed images are called  $G$ .

(2) It is required to calculate  $G'$ . For any point  $A \in G'$ , and the coordinate is  $(x_0, y_0)$ . If the positive number  $\sigma$  exists, then  $G'(x_0, y_0) = \max(G'(x, y))$ , in which

$(x - x_0)^2 + (y - y_0)^2 \leq \sigma$ . And such points will be recorded as the collection  $V$ .

(3) It is necessary to calculate  $G''$ . For any point  $A \in G''$  and the coordinate is  $(x_0, y_0)$ . If  $G''(x_0, y_0) = 0$ , then such points can be recorded as the collection  $W$ .

(4) It is required to compare  $W$  with  $V$  to obtain the point set  $T$ ,  $T = \{(x_0, y_0) \mid (x_0, y_0) \in V, \text{ and } (x_0, y_0) \in W\}$ . Then the effect of removing the “pseudo edge point” will be achieved.

Method 2: it is needed to consider the third derivative of the image as the second derivative of the real edge point and the “pseudo edge point” is zero. If the derivative is computed again, the third derivative of the real edge point will not be zero, because the “zero intersection” nature of the second derivative has ensured that its tangent at this point should be diagonal. However, the third derivation of the “pseudo edge point” must be zero, because its second derivative is the minimum. The specific steps are similar to the first method and then the edge point set will be obtained

$$T = \{(x_0, y_0) \mid (x_0, y_0) \in \{G'(x_0, y_0) = 0\} \cap \{G'''(x_0, y_0) \neq 0\}\}$$

Method 3: it is required to consider the positive and negative nature of the adjacent points around the point whose second derivative is zero. The second derivative value of the adjacent points around “pseudo edge point” shares the same sign and the second derivative value of the adjacent points around real edge point shares the different signs. After the multiplication, the one which is greater than zero is called “pseudo edge point” and the one which is less than zero is named the real edge point. As for this method, two aspects should be paid more attention to: first, which kind of range is called “adjacency”. It relies on the accuracy of CCD. The smaller the pixel is, the more the sampling points are, then the more real the “adjacency” will be; second, due to the various factors, if the multiplication of the adjacent points around certain point equals to zero, then whether these points are edge ones will be determined by more information.

## 4 Experiment and Conclusion

The edge detection results of the barcode image in digital levels before improving the LOG operator and the detection results after improving the LOG operator have been shown in Figure 3 and Figure 4. The edge points of latter are less than the ones of former, which indicates that the method of removing the “pseudo edge point” is effective.

In summary, based on the three methods of removing the “pseudo edge point” that the zero point is not always the edge point in image detection, it is required to select the appropriate LOG operator parameters. Then the edges of the barcode image of digital level will be better detected.

**Acknowledgments.** This work is supported by the special research project for the selection and the training of outstanding young teachers in Shanghai universities (Grant No.: gjd09041).

## References

1. Chen, G., Le, Y., Wu, J.: Key Algorithm of Measurement System with Digital Level. *Opto-Electronic Engineering* 35(4), 69–73 (2008)
2. Chen, G., Liu, J.: Edge Detection Methods for Gray-R Barcode Grade Rod. In: *International Conference on wireless Networks and Information Systems, WNIS 2010*, vol. 1, pp. 82–84 (December 2010)
3. Zhao, Z., Wan, J.: New method for image edge detection based on gradient and zero crossing. *Chinese Journal of Scientific Instrument* 27(8), 821–824 (2006)
4. Lv, X., Wang, H., Liu, J.: Sub-pixel edge detect technique based on LOG operator. *Journal of Baotou University of Iron and Steel Technology* 21(2), 165–168 (2002)

# New Approaches to Apply Multi-stereo Matching into 3DCG Modeling

Ningping Sun and Hiroshi Nagaoka

Kumamoto National College of Technology, Kumamoto, Japan  
sningping@kumamoto-nct.ac.jp

**Abstract.** We introduce stereo matching into three dimensional computer graphics (3DCG) modeling and expand one pair of stereo cameras to a camera network consists of a series of web cameras. We call our new approach multi-stereo matching. In this paper we present the stereo axes error correcting algorithm, approximate calculation of the stereo coefficient, and measurement method of multi-stereo. Finally we give our experimental results.

**Keywords:** Computer vision, stereo matching, error correction, 3DCG modeling.

## 1 Introduction

Stereo matching is a well-known technique for coordinate or space measurement. It is not only very easy to understand but also to program. With stereo matching we can get the exact position of a target from two stereo pictures. However the traditional stereo matching needs some strict conditions about camera, such as the axes of stereo cameras (a pair of camera with the same properties) must be kept paralleled and the height of cameras will be the same. These conditions limited the application of stereo matching.

We have introduced stereo matching into our interactive interface of 3DCG modeling because of its simplicity. We developed some algorithms and methods to relax the conditions of stereo matching and expand one pair of cameras to a series of web cameras (a network of web cameras). We call our new approach “multi-stereo matching”. Using the proposed method we can take the measurement of 3D coordinate of target inside and outside very well, and do CG modeling of object having complicated shapes as well. In this paper we present the stereo axes error correcting algorithm, approximate calculation of the stereo coefficient, and measurement method of multi-stereo. Finally we give our experimental results.

## 2 The Proposed Stereo Correcting Algorithm

The coordinate of a point  $P(x, y, z)$  can be measured by traditional stereo matching from two pictures called left image and right image taken by two cameras separately. Based on the triangulation from two views and under the conditions that the axes of the

two camera lenses must be paralleled to axis  $z$  like the red lines of Fig.1 and the height of two cameras must be the same, the coordinate of target can be computed easily.

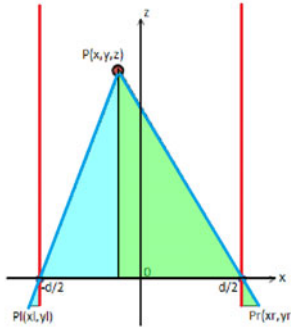


Fig. 1. Computation of traditional stereo matching

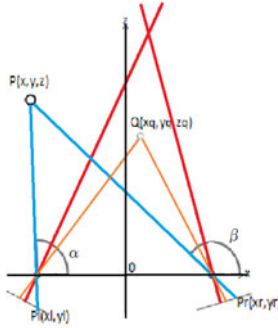
In Fig. 1 the point  $P$  is our target. For convenience, we assume that two cameras have the same focus distance  $f$ , the distance between the two cameras is  $d$ ,  $(x_l, y_l)$  denote the coordinate of  $P$  in the left image and  $(x_r, y_r)$  denote the coordinate of  $P$  in the right image, we can get  $P(x, y, z)$  by (1), (2) and (3).

$$x = \frac{d}{2} \cdot \frac{x_l + x_r}{x_l - x_r} \tag{1}$$

$$y = \frac{y_l \cdot d}{x_l - x_r} \tag{2}$$

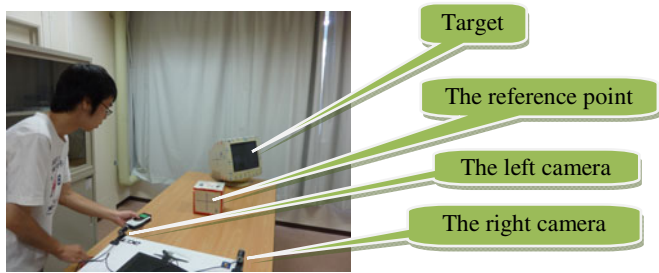
$$z = \frac{f \cdot d}{x_l - x_r} \tag{3}$$

In our experiments a number of cameras are located around of target and the two pictures taken by each pair of cameras are used for stereo matching calculation. It is impossible to compute the coordinate of target with (1), (2) and (3) if the axes of camera lens are not paralleled each other like the case of Fig.2 in which the red lines indicate the axes of two cameras. However, the axes of cameras perhaps are unparallel usually in the actual measurement unless the special stereo cameras, so that we need to relax the strict conditions of traditional stereo matching. Assume the cameras having the same height we proposed a stereo axes correction algorithm and new equations to solve the unparallelled lens axes problem.

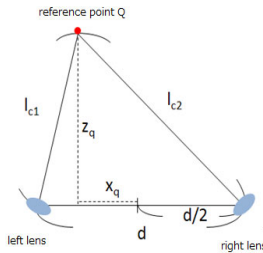


**Fig. 2.** The principle of stereo axes correction

We do know the two axes of camera lens are not paralleled but we are not sure exactly how much they incline each other. In order to find out the lens axes inclined angle we pick up a known point  $Q(x_q, y_q, z_q)$  as the reference point (Fig.3). We use a laser measure tool to get three necessary parameters: the distance between two cameras, *i.e.*,  $d$ ; the distance between the left camera and  $Q$  denoted as  $l_{c1}$  and the distance between the right camera and  $Q$  denoted as  $l_{c2}$ . Fig. 4 and (4) give the method to calculate the coordinate of point  $Q$ .



**Fig. 3.** Stereo measurement and reference point



**Fig. 4.** Measurement of the reference point

$$x_q = \frac{l_{C1}^2 - l_{C2}^2}{2d}, \quad z_q = \sqrt{l_{C1}^2 - (x_q + \frac{d}{2})^2} \quad (4)$$

Now we introduce  $Q(x_q, y_q, z_q)$  into (5) and (6) where  $x_{lq}$  is coordinate  $x$  of  $Q$  in the left image and  $x_{rq}$  is coordinate  $x$  of  $Q$  in the right image,  $\alpha$  and  $\beta$  are inclined angles of target  $P$  in the left image and the right image individual that can be expressed as the following.

$$\alpha = \text{Tan}^{-1} \left( \frac{z_q}{x_q + \frac{d}{2}} \right) + \text{Tan}^{-1} \left( \frac{x_{lq}}{f} \right) - \text{Tan}^{-1} \left( \frac{x_l}{f} \right) \quad (5)$$

$$\beta = \text{Tan}^{-1} \left( \frac{z_q}{x_q - \frac{d}{2}} \right) + \text{Tan}^{-1} \left( \frac{x_{rq}}{f} \right) - \text{Tan}^{-1} \left( \frac{x_r}{f} \right) \quad (6)$$

Therefore the coordinate of target  $P$  can be obtained by

$$x = -\frac{d}{2} \cdot \frac{\tan(\alpha) + \tan(\beta)}{\tan(\alpha) - \tan(\beta)} \quad (7)$$

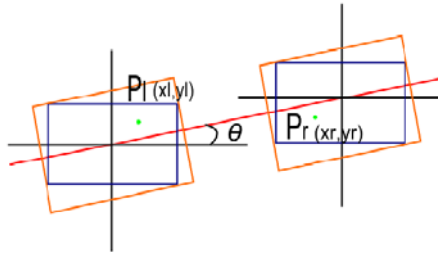
$$y = \frac{y_l \cdot \sqrt{\left(x + \frac{d}{2}\right)^2 + z^2}}{\sqrt{f^2 + x_l^2}} \quad (8)$$

$$z = \tan(\alpha) \cdot \left(x + \frac{d}{2}\right) \quad (9)$$

Having the above new equations we do not worry about whether the lens axes of cameras are paralleled or not anymore. On the other hand, in the new method the reference point plays an important role to correct the unparallelled lens axes, we should pay much attention to the measurement process shown in Fig. 3 and 4.

### 3 Correction of Different Camera Height

The proposed equations obviously should be computed under the condition of  $y_l = y_r$ . However, we cannot guarantee that cameras will be kept on the same height under any measure environment. Similarly to the situation of lens axes we are not sure the height of cameras whether is the same or not, especially when we work outside. But the images can give us some hints. For instance in the two blue images of Fig. 5 the point  $P$  has different  $y_l$  and  $y_r$  because the two blue images are taken by two cameras of different height. From our observation we found out the different height problem can be solved by rotating both of the left image and the right image to get two new dummy images that are orange ones of Fig. 5. The coordinate of point  $P$  in both of two new dummy images are changed and the new  $y_l$  is as same as the new  $y_r$ . It seems as if the new two dummy pictures had been taken by two cameras having the same height.



**Fig. 5.** Correction of camera height

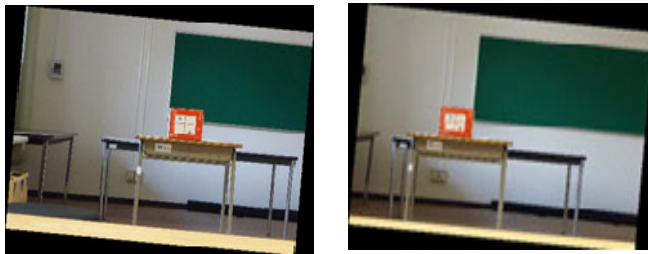
The angle of rotation can be obtained from coordinates of point  $P$  within the left image and the right image by (10).

$$\theta = \tan^{-1}\left(\frac{y_l - y_r}{x_l - x_r}\right) \tag{10}$$

Now we give an experimental example of the camera height correction in Fig. 6. The red box is our target. Because the height of the left camera and the right camera are different, the height of box in the two images are different also. After rotating we get two new images that has the same height and they can be employed in the computation of (7) (8) and (9).



(a) Before(target has different height in the left image and the right image)



(b) After (target has the same height in both images)

**Fig. 6.** Rotation of images

### 4 Calculation of Stereo Coefficient

The proposed correction algorithm and method can work well with general digital cameras or web cameras. As mentioned at the section of introduction, in order to apply multi-stereo matching into 3DCG modeling we introduce a network of web cameras to construct our sensor network rather than the special stereo cameras. The web camera has very well high-resolution and easy to be used as outside broadcasting camera of motion capture. However it is difficult to figure out the parameter  $f$ , focus coefficient of web camera, which is necessary in stereo measurement and coordinate computation. For convenience we call  $f$  stereo coefficient in our work. From our measure experiments we see that different  $d$  and unparallel axes of cameras will bring different  $x_l$  and  $x_r$ , meanwhile involved  $f$  will be different too. In the case we have to recompute  $f$  as well.

For finding a clue how to figure out stereo coefficient  $f$  we investigated the relationship among  $x$ ,  $d$ ,  $f$  in (7) furthermore. We rewrite  $\tan(\alpha)$  and  $\tan(\beta)$  of (7) as below.

$$\tan(\alpha) = \frac{\lambda_l f^2 + f(x_{lq} - x_l) + \lambda_l x_{lq} x_l}{f^2 - \lambda_l f(x_{lq} - x_l) + x_{lq} x_l} \tag{11}$$

$$\tan(\beta) = \frac{\lambda_r f^2 + f(x_{rq} - x_r) + \lambda_r x_{rq} x_r}{f^2 - \lambda_r f(x_{rq} - x_r) + x_{rq} x_r} \tag{12}$$

where  $\lambda_l = \frac{z_q}{x_q + \frac{d}{2}}$ ,  $\lambda_r = \frac{z_q}{x_q - \frac{d}{2}}$ .

It is obviously  $f$  is a non-linear parameter related to  $d$  within (7), (8) and (9). Some experimental results shown Fig. 7 tell us  $x(d, f)$  is a hyperbola.

In practice to simplify the problem we developed an alternative algorithm to approximate  $f$ . First by the method of Fig. 3, 4 and (4) we designate a known point  $U(x_u, y_u, z_u)$  as the reference point and  $U$  is not  $Q$ . Next, assume initial  $f$  as  $f_0$  and take a small positive  $df$  as variation. Apply Newton Raphson method into (7) and (13).

$$x(f) = x_u(f) - x_u = 0 \tag{13}$$

Let  $\varepsilon$  be a small positive number and start from  $f_0$ , do replace  $f_{i+1}$  with the right term of (14) repeatedly until  $x(f_i) \cdot x(f_{i+1}) < 0$ .

$$f_{i+1} = f_i - \frac{x(f_i) \cdot df}{x(f_i + df) - x(f_i)} \tag{14}$$

In order to obtain a precise  $f$  it is important and necessary to choose one pair of adequate  $f_0$  and  $df$  by through trial and error according to the results of Fig.7 because the Newton Raphson method perhaps is broken by a bad  $f_0$  or  $df$ . When recursive computations of (14) terminate at  $x(f_i) \cdot x(f_{i+1}) < 0$  we use bisection method to find  $f$  continuously by computing (15) repeatedly until  $|x(f)| < \varepsilon$ .



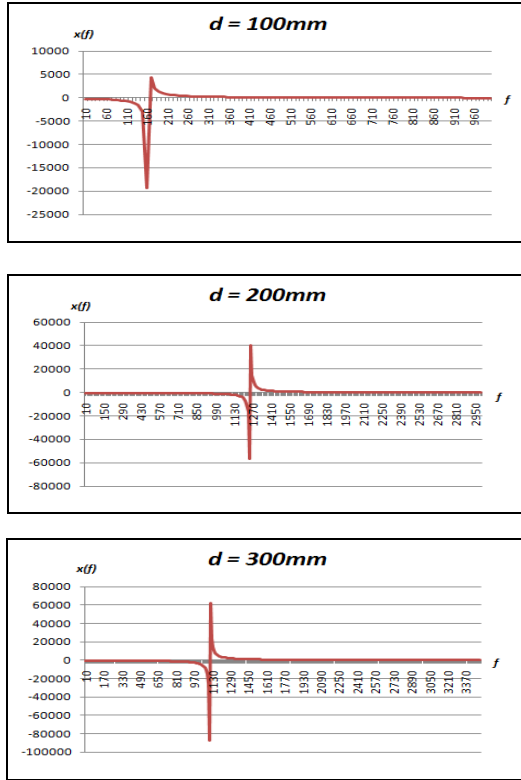


Fig. 7. Experimental results of  $x(d, f)$

$$x(f) = x_u(f) - x_u, \quad f = (f_{max} + f_{min})/2 \tag{15}$$

Now we provide some results of computing  $f$  in Table 1 where  $d$  indicates the distance between two cameras,  $f_0$  is the initial approximate value of  $f$  and  $df$  is the variation,  $i$  indicates the number of computations in which the number of Newton Raphson approximations is  $i_n$ . Responding the given  $d$  and  $x_u$  result  $f_i$  is the desired stereo coefficient. The speed of convergence of Newton Raphson method is fast but computations of (14) always fail in the nearest neighbor of  $x_u$  where  $x(f_i) \cdot x(f_{i+1}) < 0$  due to the term of numerical differentiation of (14). We combine Newton Raphson method with the bisection method in there to continue the computation in order to obtain a satisfying answer.

**Table 1.** Results of computation of  $f (\varepsilon = 0.01)$

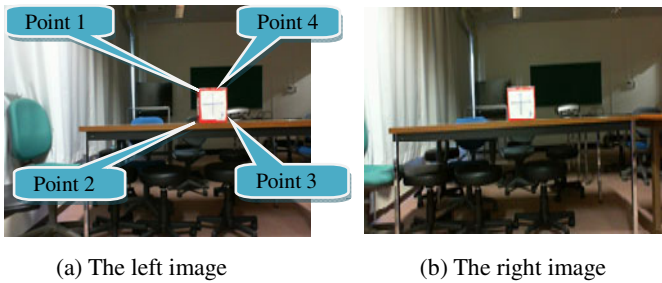
$d(mm)$	$f_0$	$df$	$i (i_n)$	$f_i$	$x_u$	$x(f_i)$
100	200	0.5	15(4)	625.12	218.99	0.007
200	1400	0.5	17(4)	2691.69	279.92	0.007
300	1300	0.5	18(5)	3269.62	359.00	0.003
400	2000	0.5	12(3)	2915.00	-90.96	0.007

## 5 DCG Modeling with Proposed Methods

It needs strict professional training and skill to model 3D models having complicated shapes by CG software tools. Apply the proposed multi-stereo matching into 3DCG could reduce the complexity of modeling and simplify design process, which provides a new kind of interactive means of computer graphics modeling and motion capture. Here we give two experimental results about our proposed correcting method and application in CG modeling.

Result 1: Examining of the proposed correcting method

In this experiment we evaluated our proposed methods including the computations of (7), (8) and (9), and stereo coefficient  $f$ .



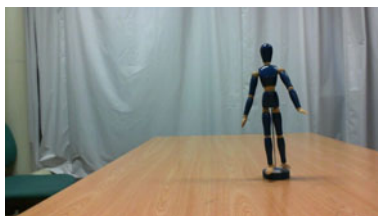
**Fig. 8.** Result 1: Examining of the proposed correcting method

**Table 2.** Results of stereo matching ( $d = 200mm, f = 2691.69$ )

$point$	$x$	$y$	$z$
1	86.34	31.05	1661.50
2	86.96	-1.24	1661.43
3	121.47	-0.62	1657.25
4	120.11	31.47	1647.11

Result 2: Motion capture with the proposed methods

In this experiment we apply stereo matching into the motion capture of animation.



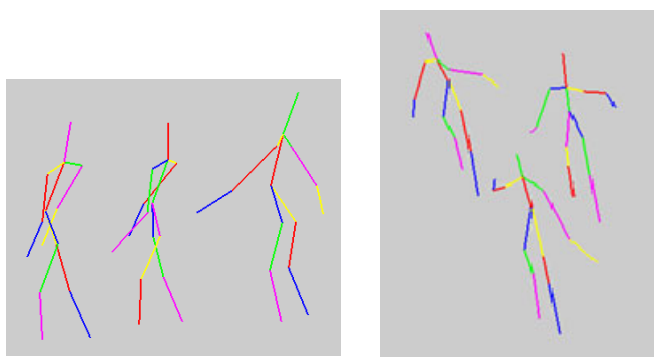
(a) Scene 1



(b) Scene 2



(c) Scene 3



(d) animation from stereo matching

**Fig. 9.** Result 2: Motion capture with the proposed methods

## 6 Conclusion and Future Work

Our approaches to apply the multi-stereo matching into 3DCG modeling can work very well and errors caused by free hand image operation can be reduced by image

processing such as edge abstraction. As the future work we shall introduce our methods into motion capture of real time.

## References

1. Forsyth, D.A., Ponce, J.: *Computer Vision: A Modern Approach*, pp. 241–281. Kyoritsu Shuppan Co. Ltd. (2003)
2. Ma, S., Zhang, Z.: *Computer Vision: Computation Theory and Algorithm Foundation*, pp. 72–94. Science Publication (1998)
3. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical Recipes in C*, pp. 251–281. Cambridge University Press (1988)

# Research on Sleep Mode in WANDAS

Li Meng<sup>1,2</sup>, Su Han<sup>2,3</sup>, Kang Yu<sup>1</sup>, and Yun Chao<sup>2,3</sup>

<sup>1</sup> Department of Automation, University of Science and Technology of China,  
Hefei, Anhui 230026, China

<sup>2</sup> Nanjing R&D Center for Broadband Wireless Communication CAS,  
Nanjing, Jiangsu 211100, China

<sup>3</sup> Shanghai Institute of Microsystem and Information Technology CAS,  
Shanghai 200000, China  
wdw820@mail.ustc.edu.cn, han.su@jushri.com

**Abstract.** This paper investigates sleep mode of MS (Mobile Subscriber) at Power Saving Class I sleep mode based on WANDAS (Wireless Access Network for Distribution Automation System). Sleep mode is a key feature introduced in WiMAX standard, which ensures minimize the MS power consumption and decrease the occupation of interface resources in serving BS (Base Station). Firstly, we propose a sleep mode control algorithm based on WANDAS, which adjusts adaptively according to the traffic rate, and based on variable sleep interval energy improved mechanism. Then we compare the proposed algorithm with the standard algorithm under four actual models which we given in this paper. Furthermore, the simulation results of the proposed algorithm show that the performance of the proposed algorithm is better than the standard algorithm in energy saving.

**Keywords:** WANDAS, WiMAX, sleep mode, energy saving.

## 1 Introduction

Stable and reliable of communication system are the key in smart grid [1]. Whether backbone network or access network must ensure that its reliability, real-time and security. Otherwise, there will be a big impact on effective, safe, stable operation and operation management of the whole system. WANDAS as an important part of smart grid [2], which development and foundation play a great role in promoting the automatic of distribution information. BS and MS operates on PMP basis in WANDAS. Since there are mass MS in WANDAS, when all of MS send distribution monitoring data to BS simultaneously, the conflicts of wireless signal is inevitable. The conflicts will lead to energy enormous waste of MS. When the MS don't transmit data to control center, we could set MS in sleep mode that could achieve the aim of save power consumption and reduce conflict.

The WiMAX standard defines a sleep mode operation [2], which can be exploited as a potential power saving mechanism. Sleep mode is a state in which the MS conducts pre-negotiated periods of absence from the BS air interface. These periods are characterized by the unavailability of MS, as observed from BS, to UL or DL

traffic. In order to minimize the MS power consumption and decrease usage of the serving BS air interface resources, the sleep mode mechanism is introduced [3].

The WiMAX standard defines three power saving classes. In Power saving Class I, the MS goes to sleep mode for  $t_{\min}$  duration [4]. Sleep windows are interleaved with listening window of fixed duration in which MS checks for any pending downlink packets at BS and in presence of traffic, MS continues to be in sleep mode with exponential increase in sleep window reaches to  $t_{\max}$ . The algorithm has one problem: MS goes to listening period frequently when the packet arrival at MS in the low arrival rate, which lead to that energy consumption too much. Therefore, we propose an improved algorithm based on Power saving Class I which can be adapted to WANDAS.

## 2 System Model

In this paper, we assume that the DL frame arrival rate and UL frame arrival to MS, following poisson distribution with rate  $\lambda_d$  and  $\lambda_u$ , the arrival time duration following

exponential distribution with the rate  $\frac{1}{\lambda_d}$  and  $\frac{1}{\lambda_u}$  respectively.

### 2.1 Sleep Interval Modeling

Let  $t_n$  denote the duration of sleep window in the  $n$ -th sleep period. The sleep window of the power saving class I mentioned in WiMAX standard is given

$$\text{by: } t_n = \begin{cases} 2^{n-1}t_{\min}, & \text{if } 2^{n-1}t_{\min} < t_{\max} \\ t_{\max}, & \text{else} \end{cases} \quad (1)$$

We propose an algorithm based on WANDAS is given by:

$$t_n = \begin{cases} \left(1 + \frac{k}{\lambda}\right) 2^{n-1}t_{\min}, & \text{if } \left(1 + \frac{k}{\lambda}\right) 2^{n-1}t_{\min} < t_{\max} \\ t_{\max}, & \text{else} \end{cases} \quad (2)$$

where  $t_{\min}$  and  $t_{\max}$  denote the minimum sleep window and the maximum sleep window, respectively. Let  $\lambda = \lambda_d + \lambda_u$  be the total arrival rate at MS.  $n$  is a positive integer.  $k$  is a constant, where  $0 < k < 0.5$ .

From equation (2), we can see that the sleep interval is dynamic adjust according to arrival rate  $\lambda$ .

### 2.2 Energy Consumption Modeling

The energy consumption by MS in sleep mode of the standard algorithm mentioned in WiMAX standard is a constant  $E_s$ , but with the passage of time,  $E_s$  would reduce along with physical equipment adjust in practical application. Therefore, we give a variable sleep period energy mechanism, the energy consumption in the n-th sleep interval meet next formula:

$$E_s(n) = \begin{cases} E_s e^{-k(t_n+L)}, & (1+\frac{k}{\lambda}) 2^{n-1} t_{\min} < t_{\max} \\ E_s e^{-k(t_{\max}+L)}, & \text{else} \end{cases} \quad (3)$$

### 3 Analytical Model

Initially, MS goes to sleep mode for  $t_{\min}$  duration. During sleep mode, if MS has any UL frame to transmit, it immediately goes to active mode. If BS has any DL frame to transmit, it not goes to active mode immediately, MS continues to be in sleep mode till end of this sleep period, then MS enters listening period. If MS receives UL frame order in listening period, then MS goes to active mode.

Using UL or DL frame transmit we determine energy consumption and average delay for all the four case [5].

**Case 1:** When UL frame arrives at MS during n-th sleep interval while there is no DL frame arrival at BS for MS.

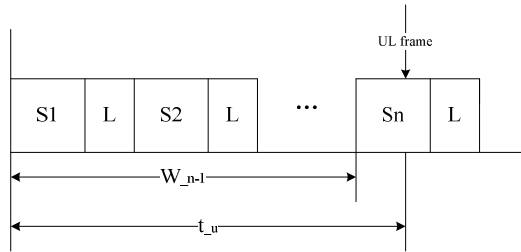


Fig. 1. UL frame arrives at MS during n-th sleep interval

Let  $S$  denote sleep interval and  $L$  is listening interval in the figure,  $t_n$  is sleep interval during n-th sleep cycle,  $t_u$  is arrival time of UL or DL frame

$$w_{n-1} < t_u < w_{n-1} + t_n$$

where  $w_n = \sum_{j=1}^n t_j + L$ .

The average time at which UL frame be present at MS is given by determining the  $t_{u\_avg}$  and is given as:

$$\begin{aligned} t_{u\_avg} &= \frac{\int_{w_{n-1}}^{w_{n-1}+t_n} x \lambda_u e^{-\lambda_u x} dx}{\int_{w_{n-1}}^{w_{n-1}+t_n} \lambda_u e^{-\lambda_u x} dx} = \frac{[-x e^{-\lambda_u x} - \frac{e^{-\lambda_u x}}{\lambda_u}]_{w_{n-1}}^{w_{n-1}+t_n}}{[e^{-\lambda_u x}]_{w_{n-1}}^{w_{n-1}+t_n}} \\ &= \frac{(w_{n-1} + \frac{1}{\lambda_u})(1 - e^{-\lambda_u t_n}) - t_n e^{-\lambda_u t_n}}{(1 - e^{-\lambda_u t_n})} \end{aligned}$$

Average energy consumption during this period:

$$E_n^1 = (t_{u\_avg} - (n-1)L)E_s + (n-1)LE_L \quad (4)$$

where  $E_s$  and  $E_L$  is energy consumption at MS during sleep interval and listening interval.

The probability that UL frame is present at MS in the n-th sleep interval is

$$\begin{aligned} P(w_{n-1} < t_u < w_{n-1} + t_n) &= \int_0^{w_{n-1}+t_n} \lambda_u e^{-\lambda_u x} dx - \int_0^{w_{n-1}} \lambda_u e^{-\lambda_u x} dx \\ &= e^{-w_{n-1}\lambda_u} [1 - e^{t_n\lambda_u}] \end{aligned}$$

The probability that no DL frame is present at BS for MS during n-th sleep interval is given as:

$$\begin{aligned} \phi_n^1 &= P(\overline{w_{n-1}} < t_u < \overline{w_{n-1} + t_n}) \cdot \\ &P(\overline{e_1}, \overline{e_2}, \dots, \overline{e_{n-1}}, \overline{e_{w_{n-1}-L}} < t < \overline{t_u}) \\ &= (e^{-w_{n-1}(\lambda_u + \lambda_d)}) [1 - e^{t_n\lambda_u}] [e^{-\lambda_d(t_u - w_{n-1} - L)}] \end{aligned} \quad (5)$$

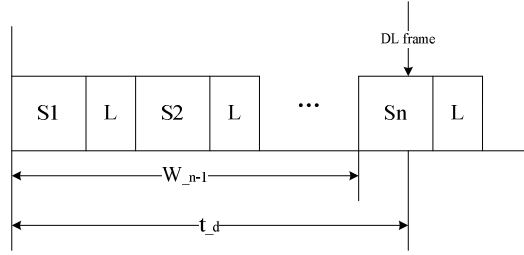
where  $e_j$  is means that DL frame is arrival during j-th sleep interval.

Because of no DL frame arrival, average delay contributed due to sleep mode:

$$D_1 = 0 \quad (6)$$

**Case 2:** When DL frame arrives at MS during n-th sleep interval while there is no UL frame arrival at BS for MS.





**Fig. 2.** DL frame arrives at MS during n-th sleep interval

Average energy consumption during this period:

$$E_n^2 = \sum_{j=1}^n t_j E_s + n L E_L \tag{7}$$

The probability that DL frame is present at MS in the n-th sleep interval is

$$\begin{aligned} P(w_{n-1} < t_u < w_{n-1} + t_n) &= \int_0^{w_{n-1} + t_n} \lambda_u e^{-\lambda_u x} dx - \int_0^{w_{n-1}} \lambda_d e^{-\lambda_d x} dx \\ &= e^{-w_{n-1} \lambda_d} [1 - e^{t_n \lambda_d}] \end{aligned}$$

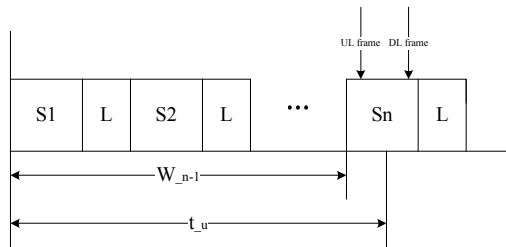
The probability that no UL frame is present at BS for MS during  $n^{th}$  sleep interval is given as:

$$\begin{aligned} \phi_n^2 &= P(t_d > w_{n-1} + t_n) \cdot \overline{P(e_1, e_2, \dots, e_{n-1}, e_n)} \\ &= e^{t_n \lambda_d} [1 - e^{-\lambda_d (t_n + L)}] \end{aligned} \tag{8}$$

Because of the MS in sleep mode, the DL frame can not transmit in time, average delay time contributed due to sleep mode is given as :

$$D_2 = \phi_n^2 \left( \frac{t_n + L}{2} \right) \tag{9}$$

**Case 3:** When UL frame arrives at MS during n-th sleep interval while there is DL frame arrival at BS for MS.



**Fig. 3.** UL frame arrives at MS during n-th sleep interval while there is DL frame arrival at BS for MS

The average time at which UL frame be present at MS is given by determining the  $t_{u\_avg}$  and is given as:

$$\begin{aligned}
 t_{u\_avg} &= \frac{\int_{w_{n-1}}^{w_{n-1}+t_n} x\lambda_u e^{-\lambda_u x} dx}{\int_{w_{n-1}}^{w_{n-1}+t_n} \lambda_u e^{-\lambda_u x} dx} = \frac{[-xe^{-\lambda_u x} - \frac{e^{-\lambda_u x}}{\lambda_u}]_{w_{n-1}}^{w_{n-1}+t_n}}{[e^{-\lambda_u x}]_{w_{n-1}}^{w_{n-1}+t_n}} \\
 &= \frac{(w_{n-1} + \frac{1}{\lambda_u})(1 - e^{-\lambda_u t_n}) - t_n e^{-\lambda_u t_n}}{(1 - e^{-\lambda_u t_n})}
 \end{aligned}$$

Average energy consumption during this period:

$$E_n^3 = (t_{u\_avg} - (n-1)L)E_s + (n-1)LE_L \quad (10)$$

where  $E_s$  and  $E_L$  is energy consumption at MS during sleep interval and listening interval.

The probability that UL frame is present at MS in the n-th sleep interval is

$$\begin{aligned}
 &P(w_{n-1} < t_u < w_{n-1} + t_n) \\
 &= \int_0^{w_{n-1}+t_n} \lambda_u e^{-\lambda_u x} dx - \int_0^{w_{n-1}} \lambda_u e^{-\lambda_u x} dx \\
 &= e^{-w_{n-1}\lambda_u} [1 - e^{t_n\lambda_u}]
 \end{aligned}$$

The probability that no DL frame is present at BS for MS during  $n^{th}$  sleep interval is given as:

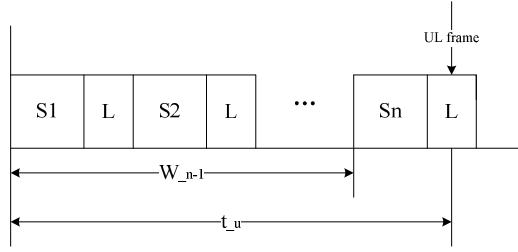
$$\begin{aligned}
 \phi_n^3 &= P(w_{n-1} < t_u < w_{n-1} + t_n) \cdot \\
 &P(\overline{e_1}, \overline{e_2}, \dots, \overline{e_{n-1}}, e_{w_{n-1}-L} < t < t_u) \\
 &= (e^{-w_{n-1}(\lambda_u + \lambda_d)}) [1 - e^{t_n\lambda_u}] [1 - e^{-\lambda_d(t_u - w_{n-1} - L)}]
 \end{aligned} \quad (11)$$

where  $e_j$  is means that DL frame is arrival during j-th sleep interval.

Because of the MS in sleep mode, the DL frame can not transmit in time, average delay time contributed due to sleep mode is given as:

$$D_3 = \phi_n^3 \left( \frac{t_{u\_avg} - w_{n-1} - L}{2} \right) \quad (12)$$

**Case 4:** When UL frame arrives at MS during n-th listening interval while there is no DL frame arrival at BS for MS.



**Fig. 4.** UL frame arrives during n-th listening interval

The average time at which UL frame be present at MS is given by determining the  $t_{u\_avg}$  and is given as:

$$\begin{aligned}
 t_{u\_avg} &= \frac{\int_{w_n-L}^{w_n} x \lambda_u e^{-\lambda_u x} dx}{\int_{w_n-L}^{w_n} \lambda_u e^{-\lambda_u x} dx} = \frac{[-x e^{-\lambda_u x} - \frac{e^{-\lambda_u x}}{\lambda_u}]_{w_n-L}^{w_n}}{[e^{-\lambda_u x}]_{w_n-L}^{w_n}} \\
 &= \frac{(w_n + \frac{1}{\lambda_u})(e^{-\lambda_u L} - 1) - L e^{-\lambda_u L}}{(e^{-\lambda_u L} - 1)}
 \end{aligned}$$

Average energy consumption during this period:

$$E_n^4 = \sum_{j=1}^n t_j E_s + \left[ t_{u\_avg} - \sum_{j=1}^n t_j \right] E_L \quad (13)$$

The probability that no DL frame is present at BS for MS during  $n^{th}$  listening interval is given as:

$$\begin{aligned}
 \phi_n^4 &= P(w_{n-1} < t_u < w_{n-1} + t_n) \cdot P(\overline{e_1}, \overline{e_2}, \dots, \overline{e_{n-1}}, \overline{e_n}) \\
 &= e^{-w_n(\lambda_u + \lambda_d)L} [e^{\lambda_u L} - 1]
 \end{aligned} \quad (14)$$

where  $e_j$  is means that DL frame is arrival during j-th sleep interval.

Because of no DL frame arrival, average delay contributed due to sleep mode:

$$D_4 = 0 \quad (15)$$

Base on the four actual models we introduced, the average energy consumed is given by

$$E = \sum_{n=1}^{+\infty} (\phi_n^1 E_n^1 + \phi_n^2 E_n^2 + \phi_n^3 E_n^3 + \phi_n^4 E_n^4) \quad (16)$$

The total average delay is given by

$$D = \sum_{n=1}^{+\infty} D_2 + D_3 \tag{17}$$

### 4 Results and Discussion

In this section, we have presented the analytical as well as simulation results and the parameters used to evaluate proposed algorithm. The simulation is achieved using Matlab. Based on the four actual models we presented in the section 3, we compared the sleep mode of the proposed algorithm with the standard algorithm mention in WiMAX. In this research, the simulation parameters set as :

$$k = 0.1, L = 1, t_{\min} = 1, t_{\max} = 1024; E_s = 50 \text{ mW}, E_L = 150 \text{ mW};$$

$$C = \frac{\lambda_u}{\lambda_d}, \lambda = \lambda_d + \lambda_u, 0.025 < \lambda < 0.5.$$

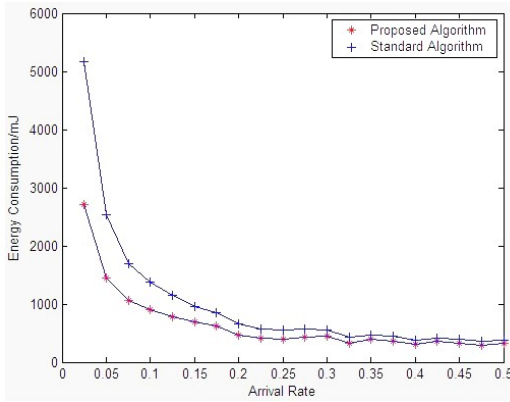


Fig. 5. Comparison of energy consumption

Fig5 shows average energy consumption by MS for standard algorithm and proposed algorithm with respect to mean arrival rate  $\lambda$ . It has been observed that the proposed algorithm reduces energy consumption significantly in low arrival rate. Significant reduction is observed because the proposed algorithm reduce the number of MS enter listening period in low arrival rate, and it ensures that the energy consumption during sleep period be reduced as time goes on based on variable sleep period energy scheduling mechanism.

Fig6 presents average time delay contributed due to sleep mode by MS for standard algorithm and proposed algorithm with respect to mean arrival rate  $\lambda$ . This study shows that the proposed algorithm time delay increase slightly at low arrival rate. The main reason is that the sleep interval is dynamic adjust according to arrival

rate  $\lambda$  in the proposed algorithm, this change lead to lager time delay at law arrival rate, but the time delay will follows the same tread for both algorithms. Since MS equipment for the energy requirements than the requirements of real-time data transmission, the proposed algorithm can reduce power consumption if the time delay is acceptable.

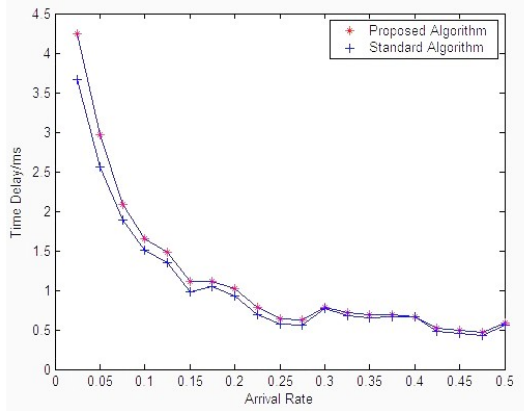


Fig. 6. Comparison of time delay

## 5 Conclusion

This paper proposed an algorithm that can be apply to wireless access network for Distribution Automation System based on the WiMAX standard sleep mode. On one hand the algorithm adjusts adaptively the sleep mode interval according to the arrival rate, on the other hand definition sleep period energy variable. Then we give four actual models according to UL or DL frame arrival. Finally, we compared the standard algorithm mentioned in WiMAX standard with the proposed algorithm under four actual models which we given in this paper, the simulation results show that the performance of the proposed algorithm in low arrival rate is better than the standard algorithm in energy saving.

**Acknowledgments.** This work was supported by Nanjing demonstration project of Internet of Things: Intelligent network monitoring System for Smart Grid and Reserve project of Chinese Academy of Sciences: Development of Multi-user scheduling system for WANDAS in Smart Grid.

To whom all correspondences should be addressed (Li Meng), E-mail: wdw820@mail.ustc.edu.cn

## References

1. Moslehi, K.: A reliability perspective of the smart grid. *IEEE Trans. on Smart Grid*, 57–64 (2010)
2. IEEE, IEEE 802.16 medium access control and service provisioning. *IEEE Trans. on Automatic Control* 8(3) (2004)
3. Dong, G., Zheng, C., Zhang, H., Dai, J.: Power Saving Class I Sleep Mode in IEEE802.16e System. In: *The 9th International Conference on Advanced Communication Technology*, pp. 1487–1491 (2007)
4. Jung, W.J., Ki, H.J., Lee, T.-J., Chung, M.Y.: Adaptive sleep mode algorithm in IEEE 802.16e. In: *Proceedings of Asia-Pacific Conference on Communications*, pp. 483–486 (2007)
5. Sanghvi, K., Jain, P.K., Lele, A., Das, D.: Adaptive waiting time threshold estimation algorithm for power saving in sleep mode of IEEE 802.16e. In: *IEEE International Conference on Communications*, pp. 649–653 (2007)

# Design and Realization of Integrated Ultrasound Transmitting Circuit

Hao Liu, Changyun Miao, Yongmao Feng, and Feng Rong

School of Electronic and Information Engineering,  
Tianjin Polytechnic University, Tianjin, 300160, China  
xhliuhao@sohu.com, miaochangyun@tjpu.edu.cn,  
fylouis@163.com, shusheng667@163.com

**Abstract.** The current ultrasound transmitting circuit has many disadvantages like large size, low digital standard, bad flexibility of adjustment and so on. A new integrated ultrasound transmitting circuit is developed. The circuit use DDS technique based on FPGA to control DC-DC boost circuit and pulse exciting circuit synchronously. Negative sharp pulses were generated which used to excite ultrasonic transducer. The results of experiment show that the transmitting circuit was stable and reliable. Compared with the current ultrasonic transmitting circuit, it has many advantages such as simple hardware structure, abundant digital port, strong expansibility and so on. And it can be adaptable to the demands of miniaturization and digitization for ultrasonic testing device.

**Keywords:** Ultrasound Inspection, DDS (Direct Digital Synthesizer), DC-DC Boost, Pulse Exciting.

## 1 Introduction

Ultrasonic inspection is widely used in fields of modern technology, and it plays an important role in our national economy. There are some weaknesses existing in present ultrasonic testing apparatus such as large size, low detection precision. Especially for the ultrasound transmitting circuit which seriously affects the precision and speed of detection, there may often be some disadvantages just as bad flexibility of adjustment, short of digital interface and so on [1]. The present ultrasound transmitting circuit may not adapt for the development trend of digital ultrasonic device.

This paper proposed a kind of integrated ultrasound transmitting circuit on basis of careful research for ultrasound transmitting. The new ultrasound transmitting circuit has many advantages such as simple structure, compact size and convenient use compared with the old.

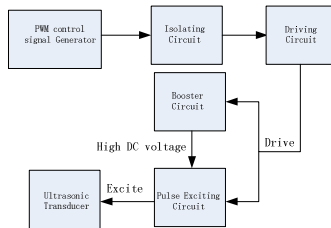
## 2 Scheme of Integrated Ultrasonic Transmitting Circuit

Ultrasonic Inspection uses the piezoelectric effect of transducer to generate and receive ultrasonic waves. When some form of voltage is forced to transducer, the

transducer will generate ultrasonic signals. On the contrary, when the transducer receives ultrasonic waves, it will convert the ultrasonic signals to voltage signals. The voltage which excites the ultrasonic transducer is generated by charging and discharging of capacitor over high DC voltage [2].

Usual ultrasound transmitting circuit is composed of high DC voltage module and exciting module independently. The high DC voltage module can be divided into two kinds [3]. The first kind uses the principle of DC inverter, and it needs transformer to convert voltage. So the size of this module is too large. The other uses DC-DC power module to complete voltage converting. Although the size of this kind was a little less, it cost much more money. Besides, the DC-DC power module sometimes can not suit your different kinds of needs. The signal of exciting module is often generated by single-chip or ARM, and it has poorer flexibility and low precision in frequency and duty ratio compared with FPGA.

The basic structure of integrated ultrasound transmitting circuit is shown in Fig 1. It is composed of PWM control signal generator, isolating circuit, driving circuit, booster circuit and pulse exciting circuit. PWM control signal generator is designed by FPGA. It supplies two output signals to control booster circuit and pulse exciting circuit synchronously. Isolating circuit can prevent interfere from entering into the digital circuit. Driving circuit is used to enhance the load capacity of PWM signal. Booster circuit can convert 12V low-voltage DC power to high DC voltage. Pulse exciting circuit generates right pulses to exciting the transducer.



**Fig. 1.** Basic structure of integrated ultrasound transmitting circuit

### 3 Design of Integrated Ultrasonic Transmitting Circuit

#### Design of PWM Control Signal Generator

PWM control signal generator needs to produce two different signals to control booster circuit and pulse exciting circuit separately. DDS technology based on FPGA can be used to generate adjustable PWM waveforms [4]. DDS is advance digital frequencies synthesize technology. It features high frequency resolution ratio, low phase noise and short-frequency switching time, and is liable to realize flexible digital modulation with high accuracy.



The main body structure of PWM control signal generator is composed of DDS circuit, counter and comparator. DDS circuit is used to generate frequency trigger signal, counter and comparator are used to set up the width of the signal can be revised by on-line programming.

The schematic and timing simulation of PWM control generator is shown in Fig 2.

### Design of Isolating and Driving Circuits

Isolating and driving circuits are shown in Fig.3. HCPL2630 is a high speed-10Mbit/s logic gate optocoupler. The device can isolate the control system from HV-power, and improve the control system reliability. Besides, the dual independent channels can realize two different signals' isolation. UCC27324 which is produced by TEXAS INSTRUMENTS is a dual 4-A peak high speed low-side power MOSFET driver. It provides driving power to signal from FPGA. The Enable port of it can reduce the power by time-sequence control.

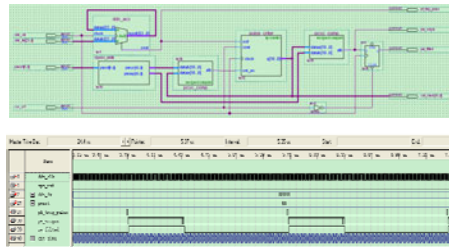


Fig. 2. The whole Structure and Timing Simulation of PWM Control Circuit

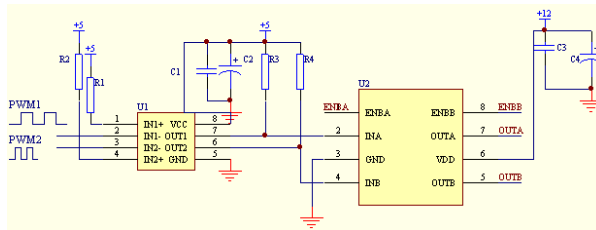
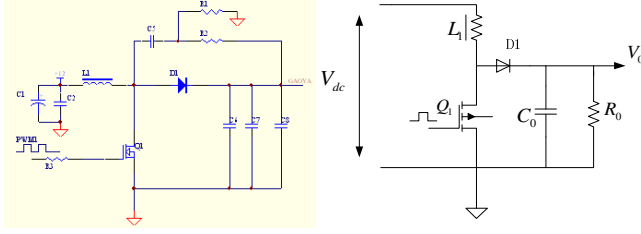


Fig. 3. Isolating and driving circuit

### Design and Analysis of DC-DC Booster Circuit

DC-DC booster circuit is shown in Fig.4. It is composed of inductor, high-voltage switch, Diode rectifier and capacities. The key device in this circuit is the high-voltage switch. The chip model is IPP90R500C3 which is produced by Infineon. IPP90R500C3 belongs to CoLMOS Power Transistor [5]. The voltage between drain and source can reach to 900V and the continuous drain current can reach to 11A. We can configure the parameter of the booster circuit by simplify the circuit which is also shown in Fig.4.



**Fig. 4.** DC-DC booster circuit and its simplified model

The transducer in the whole circuit is capacitive resistance, and the current running through it is low. As a result, the DC-DC booster ratio can be raised while the efficiency of power supply needs to be reduced. Assuming the energy of inductor can be transformed to the output completely,  $T_{ON}$ ,  $T_{OFF}$  and  $T_s$  are supposed as the switch ON- time, OFF-time and cycle time.

We have seen that when Q1 turns “on”, the current ramps up linearly in L1 to a peak value:

$$\Delta I_{L(+)} = \frac{V_{dc}}{L_1} T_{ON} \tag{1}$$

When Q1 turns “off”, the current in L1 is:

$$\Delta I_{L(-)} = \frac{V_o - V_{dc}}{L_1} T_{OFF} \tag{2}$$

In steady state, the variation of “on” is equal to “off”:

$$\frac{V_{dc}}{L_1} T_{ON} = \frac{V_o - V_{dc}}{L_1} T_{OFF} \tag{3}$$

The input power at “on” time and “off” time is:

$$P_{ON} = V_{dc}^2 T_{ON} / L_1 \tag{4}$$

$$P_{OFF} = V_{dc} (V_o - V_{dc}) T_{OFF} / L_1 \tag{5}$$

We integrate  $P_{ON}$  and  $P_{OFF}$  between  $0 \sim T_{ON}$  and  $0 \sim T_{OFF}$  separately, and then take the average value:

$$P_{AVG} = \frac{1}{T_{ON} + T_{OFF}} \left( \int_0^{T_{ON}} \frac{V_{dc}^2}{L_1} t dt + \int_0^{T_{OFF}} \frac{V_{dc} (V_o - V_{dc})}{L_1} t dt \right) \tag{6}$$

Through (3) and (6) we obtain:

$$P_{AVG} = \frac{1}{T_{ON} + T_{OFF}} \left( \frac{V_{dc}^2}{2L_1} T_{ON}^2 + \frac{V_{dc}^2}{2L_1} \cdot \frac{V_{dc}}{V_o - V_{dc}} T_{ON}^2 \right) \tag{7}$$

Assuming there is no loss in booster circuit, we know the input power is equal to the output power:

$$\frac{1}{T_{ON} + T_{OFF}} \left( \frac{V_{dc}^2}{2L_1} T_{ON}^2 + \frac{V_{dc}^2}{2L_1} \bullet \frac{V_{dc}}{V_o - V_{dc}} T_{ON}^2 \right) = V_o I_o \tag{8}$$

We can get the expression of output voltage after simplified:

$$V_o = V_{dc} + \frac{(V_{dc} D)^2}{2L_1 I_o f_s} \tag{9}$$

Where  $D = \frac{T_{ON}}{T_s}$  and  $f_s = \frac{1}{T_s}$ .

We know there is relationship between output voltage and inductor  $L_1$ , load current  $I_o$ , duty ratio  $D$ , operation frequency  $f_s$ . When the input voltage is a definite value, in order to obtain higher voltage and better ripple, the duty cycles must to be as long as possible. The whole booster circuit should work under condition of high inductance, low current and low frequency.

### Design and Analysis of Pulse Exciting Circuit

Pulse exciting circuit is show in Fig.6. It is composed of a charging resistor, a switch transistor, a high-voltage capacity, damping resistor and fast switching diodes. We use MTP4N85 which is produced by Freescale as the switch transistor. It features high voltage endurance and fast switching time. The fast switching diodes is used to restrain the positive voltage.

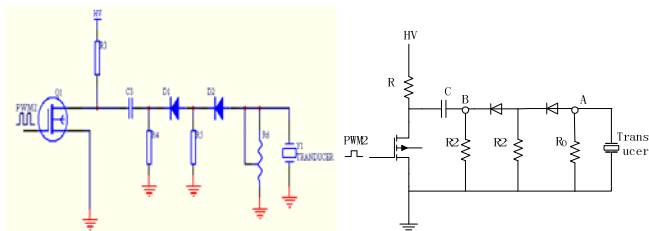
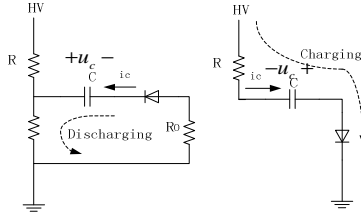


Fig. 5. Pulse exciting circuit and its simplified model

The simplified pulse exciting circuit is also shown in Fig.4. When PWM2 signal is high, the switch transistor is open, and the circuit is in status of discharge. On the contrary, when PWM2 signal is low, the switch transistor is closed, and the circuit is in status of charge. The equivalent circuit models of charge and discharge are shown in Fig.6, where R1 is  $R_{dson}$  of the switching transistor. The two resistors signed R2 are parallel to R0 during the discharge. The two much higher resistors can be ignored during this time. While during the charge, the electric potential of A was restrained by diodes. We can figure the potential of B up to determine the parameters of other devices.

We assumes  $V$  is the value of high voltage,  $R$  is the value of charging resistor,  $C$  is the value of capacity,  $R_0$  is the value of damping resistor,  $R_1$  is the value of  $R_{dson}$  of Q1,  $R_2$  is value of shunt resistor.



**Fig. 6.** Model of pulse exciting circuit

The equation of discharge circuit described by KVL (Kirchhoff's voltage law) is:

$$\frac{R_1}{R + R_1}V + i_c R_1 + i_c R_0 - u_c = 0 \tag{10}$$

The current of discharge circuit:  $i_c = -C \frac{du_c}{dt}$ , substituting for  $i_c$ , in (10) we get

$$C(R_0 + R_1) \frac{du_c}{dt} + u_c = \frac{R_1}{R + R_1}V \tag{11}$$

This is a typical differential equation for electric. Assuming  $C(R_0 + R_1) = \tau$ ,  $R_1 V / (R + R_1) = u$ , the equation can be simplified as below:

$$\frac{du_c}{dt} + \frac{1}{\tau} u_c = \frac{u}{\tau} \tag{12}$$

Substituting for the initial condition of  $t=0$ ,  $u=V$ , the result of (12) can be figure out as below:

$$u_c = \frac{R_1}{R + R_1}V + \frac{R}{R + R_1}V e^{-\frac{t}{\tau}} \tag{13}$$

$$i_c = \frac{R}{R + R_1}V \frac{1}{R_0 + R_1} e^{-\frac{t}{\tau}} \tag{14}$$

The equation of charge circuit described by KVL (Kirchhoff's voltage law) is:

$$(R + R_2)i + u_c = V \tag{15}$$

Similarly, we can figure out the result of (16) as below:

$$u_c = V \left( 1 - \frac{R}{R + R_1} e^{-\frac{t}{\tau_1}} \right) \tag{16}$$

$$i_c = \frac{R}{R + R_1}V \frac{1}{R} e^{-\frac{t}{\tau_1}} \tag{17}$$

According to the mathematic deduction above, at the beginning of discharge, the voltage which is exciting transducer, that is, the electric potential of A is

$u_A = -\frac{R_0RV}{(R + R_1)(R_0 + R_1)}$ . And it decreased to zero at a very short time. Because

resistance of R is much bigger than that of R1 and R0, the signals exciting the ultrasonic transducer are negative voltage pulses whose amplitude reaches to V.

While at the beginning of charge, the electric potential of B is  $u_B = \frac{R_2RV}{(R + R_1)(R + R_2)}$ ,

here diodes are introduced to restrain positive voltage to affect ultrasonic transducer. The waveform is shown in Fig.7.

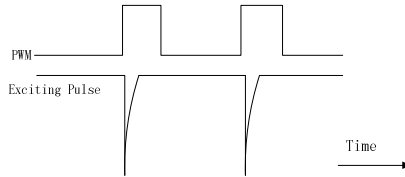


Fig. 7. Exciting pulse waveform

### 4 Testing Results

The photo of the integrated ultrasound transmitting circuit and the test result are shown in Fig.8. We will find the parameter of the ultrasound transmitting circuit meet standard through the waveforms shown in oscillograph. The frequency of PWM1 is 5.1 kHz, and the duty ratio is 1370/1550. The frequency of PWM2 is 200Hz, and the duty ratio is 10%. We will see the booster circuit convert the 12V dc voltage to over200V, and pulse exciting circuit generates negative pulse to excite the transducer.

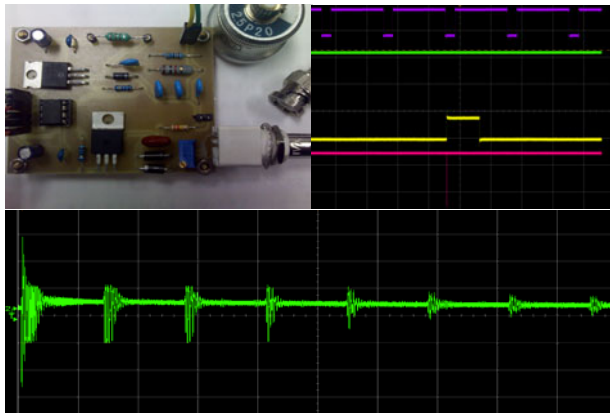


Fig. 8. Photos and test result

## 5 Conclusions

To improve the present ultrasound transmitting circuit, the paper proposed an integrated ultrasound transmitting circuit for ultrasonic inspections. The circuit has a smaller size, better transmitting performance and lower power cost. When we use it to generate ultrasound signals, EMI (Electro-Magnetic Interference) needs to be paid attentions to. The parameters of it are well through testing. Using this design of transmitting circuit, the size of ultrasound test device can be decreased obviously. Also, stronger digital controllability is especially appropriate for automation device to build multichannel ultrasound inspection system.

## References

1. Feng, H., Xiao, D., Xu, C., Zhou, S.: Device of Circuit for Transmitting/Receiving Ultrasonic Pulse. *Instrument Technique and Sensor* (11), 30–32 (2003)
2. Cai, R.: *Research On High-integration Ultrasonic Phased Array Inspection System Related Technology*, vol. 5. Tianjin University (2009)
3. Yi, L., Song, S., Ding, L., Xiao, J.: The Transmitting Circuit Design for Ultrasonic Based on Low Power. *Chinese Journal of Scientific Instrument* 04, 557–560 (2008)
4. Jia, K., Zhao, Y., Lv, Y.: Design and implementation DDS based on FPGA. *IEEE Circuits, Communications and System* 8(2), 121–124 (2010)
5. Lorenz, L., Deboy, G., Knapp, A., Marz, M.: COOL-MOSTM—a new milestone in high voltage power MOS. In: *Proc. ISPSD*, pp. 3–10 (1999)

# Special Way of Optical Communications the Technique of Superluminal Communications

Zi Hua Zhang

P. O. Box 171 Bupt, Beijing 100876, China  
Apt. 22-441, No.10 Xi Tu Cheng Lu, Haidian distract, Beijing 100876 China  
zhangzihua01@126.com, 86-10-62281482

**Abstract.** We discussed new way---superluminal communication to enhance the information transmission capacity. Proved whether  $u > c$ , or  $u < c$  the coded pulses keep in order as long as the observation outside the medium, In this paper we also responded the challenge facing to the superluminal transmission of information, in our opinion the predication of Einstein, no body (include information) can move faster than  $c$ , is wrong. In this paper we introduced our research result and some problems should be solved in future.

**Keywords:** superluminal communications, information superluminal transmission, Causality.

## 1 Introduction

We are living in an Information era, accompanying with the development of INTERNET, the information that need to exchange increate rapidly, therefore we still need to look for new technique of information transmission with higher speed. Any communication system is consists of two parts, one is the signal treatment and other is the transmission of signal, the former also contains the generation and detection of signal. Of cause, a super high speed communication system also need a suitable techniques of the generation and detection of signal, at present they are almost Opto-electronic, Optical or Quantum, and belong to front problem of informatics. Here we only discuss the problem of signal transmission. In digital communications for wireless or optical, the information transmission is by virtue of coded pulses the transmission speed of information converts to the problem of transmission speed of pulses that is determined by refractive index and dispersion of medium. For wireless the speed is about  $c$ ; but for fiber communication, the speed is equal to  $c/n_1$ , where  $n_1$  is the refractive index of fiber core. Since the short pulse suffers a serious distortion easily [1] so that the signal pulse can't be too narrow, moreover the transmission speed of pulses is restricted, for getting a high transmission rate we have to employ dense wavelength division multiplex (DWDM) technique. Right now, The Channel number is as reported as high as 256, all of them transmit in one fiber at a time, complicate the techniques of compensation of non-linearity and dispersion of the fiber as well as the multiplex and de-multiplexing, meanwhile make the communication

system very complex and expensive. Even restrict the property of communication system. We can parable the communication just as the traffic, the capacity of communication system will direct ratio to the speed of signal transmission. And we can say that for a high speed system the faster information current, the higher Transmission rate and the less channel number needed. At present, a lot of superluminal propagation phenomena of the light pulses have been observed [2], even the velocity of light pulses has a negative value (about  $-310c$ ) [3], it means the transport velocity of light pulses in the Cs vapor can faster than positive infinite and time advance happen, if these light pulses be employed to transmit information, one channel is enough to transport several T bits signal, the system will be simplified, and the problems appear in multi channel also can be avoided, even the transport time of information can be shorted too. Here we shall discuss the techniques of superluminal transmission of information, and introduce our some research results.

## 2 Superluminal Propagation of Light Pulse and Superluminal Optical Communication

### 2.1 Group Velocity and Anomalous Dispersion

In Optics, light is the electro-magnetic wave, with wavelength lies in a certain region, for visible light is  $0.4-0.75\mu$ . Light pulse contain a series of frequency components, the pulse peak propagates with “group velocity”  $v_g$ , but the single monochromatic wave propagates with the phase velocity  $v_p$ . In general, the  $v_g$  and  $v_p$  are determined by the electro-magnetic parameters of medium and both of them can larger than  $c$ . In the case of the linear effect be considered only, the transmission property of medium is determined by its polarizability  $\chi(v)$ . We have [1]

$$\begin{aligned} n(v) &= 1 + \text{Re}[\chi(v)] \\ \alpha(v) &= \frac{2\pi v}{nc} \text{Im}[\chi(v)] \end{aligned} \quad (1)$$

Where  $\text{Re}$  represent the real part and  $\text{Im}$  is imaginary part of  $\chi(v)$  respectively,  $\alpha$  represents absorption coefficient when  $\alpha > 0$ , or gain coefficient when  $\alpha < 0$  and when  $\alpha = 0$ , the medium is transparent, the dispersion relationship of medium can be written as

$$k(v) = 2\pi \frac{n(v)v}{c} \quad (2)$$

According to the definition, the phase velocity and group velocity can be written as

$$v_p = \frac{\omega}{k} = \frac{c}{n}, \text{ and } v_g = \frac{d\omega}{dk} = \frac{c}{n_g} \quad (3)$$



From equations (2) and (3), we can get

$$n_g = n + \omega \frac{dn}{d\omega} \quad (4)$$

And 
$$v_g = v_p - \lambda \frac{dv_p}{d\lambda} \quad (4')$$

We have proved that the group velocity just is the Velocity of pulse peak in theory [4].  $dn/d\omega$  is dispersion of medium. Since  $n \approx 1$  in air, for  $v_g > c$ , we need  $dn/d\omega < 0$ , it means that the light pulse have to propagate in anomalous dispersion medium. In normal dispersion region, since  $dn/d\omega > 0$ , so that we have  $n_g > n$ , in this case the group velocity always less than phase velocity  $v_p$ , Sometimes  $n_g$  can be very larger make the group velocity reducing to 8m/s [5]. In anomalous dispersion region, although the value of  $dn/d\omega$  is very small, but the angle frequency  $\omega$  is very larger, so there are following cases:

- 1,  $1 < n_g < n$ ; since in such case  $n_g > 1$ , we always have  $v_g < c$ ;
- 2,  $0 < n_g < 1$ ; in such case we always have  $\infty > v_g > c$ ;
- 3,  $n_g = 0$ ; in such case we have  $v_g = \infty$ ;
- 4,  $n_g < 0$ ; the value of  $v_g$  is negative.

It shows when the light pulses propagates in anomalous dispersion region, the refractive index of group velocity can less than 1, even has a negative value. In other words, the pulse peak can move faster than  $c$ , even group velocity is negative. Our research shows that the negative value of group velocity is faster than the velocity with positive infinite value, and the time advance happen [4].

## 2.2 The Relationship between Propagation of Light Pulse and Information Transmission

We have proved that the superluminal propagation of light pulses is caused by the interaction between light and matter [4], and also showed that the group velocity  $v_g$  is the velocity of the pulse peak and can larger than  $c$ , even negative. Since these pulses can be detected completely and still is the concentration of electric field or photons, of cause they can be employed to transfer the Information, means that the velocity of information transmission also can faster than  $c$ , that conflict with the special Relativity. Since Einstein had predicated no body (include the information) can move faster than  $c$ , otherwise the inversion of causality will happen [5,6]. Sommerfeld and Brillouin also had pointed out that, since the causality need the velocity of information transmission not faster than  $c$ , so that when group velocity is larger than  $c$ , and it no longer represents the velocity of a signal or of energy transport, the signal isn't the light itself with move with a group velocity [7], they further defined the velocity of signal as "frontal velocity" instead of the signal velocity which cause misunderstanding easily, according to their definition the front contains a infinite bandwidth that can't realize. In the studies of problem of information transmission, they definite five velocities, especially they defined the velocity of energy transport  $v_e$  again and demonstrated that the velocity of energy transport always less than light

speed in vacuum  $c$ . Since that, some persons consider the superluminal pulse is virtual and contains no energy, so that these superluminal pulses can't transfer information. Is this point of view correct? In fact, the superluminal pulses had been detected. For showing the superluminal communication is possible. We must answer these questions in theory.

At first, we have proved theoretically both of the assertion and the method of deduction of the velocity of energy transport can't larger than  $c$  is wrong [8]. Because of the inner energy of the transmission medium not take part in the energy transport, for a monochromatic wave the velocity of energy transport just equal to the phase velocity, meanwhile for a light pulse the concept of the velocity of energy transport is ambiguous in a dispersive medium, when the distortion of pulse isn't symmetry, we give a definition of the velocity of energy transport difficulty. Actually, curve  $c/v_c$  in Brillouin's figure [7] only corresponds to the contour of the spectral line.

Secondly, for the contradiction between superluminal transmission of light pulses and the Special Relativity, we think that Einstein made a mistake, the SR has been misread, we have written a paper to discuss this problem [note 1], here only introduce its content in next section brief. We have demonstrated when  $c < v_g < \infty$  no violation of causality, the signals keep in order, but when  $v_g < 0$  the time advance of output pulses happen due to the formula  $L = v_g \times t$  exist, we think in that time, the events obey a convert causality [7], but after the signals pass through the transmission medium the signals also keep in order, so that as long as we make detection of signal outside the transmission medium, the change of signal velocity has no influence on the order of signals. Wood had observed the light speed is larger than  $c$  in the Na vapor nearby the yellow spectral line [8] in 1904. For forcing the superluminal group velocity fit the restriction of SR, at present many scientists still consider that the light pulses (signal) with a transiting velocity larger than  $c$  is virtual, the superluminal transmission of light pulses is caused by visual error. They can't interpret why these virtual light pulses can cause the opto-electronic effect and can be detected? In our opinion, these pulses still represent a concentration of electric field or Photons, and the superluminal light pulse is real and detectable so that they can be employed to transmit information certainly. As we known, In digital communication the information transmission is by virtue of a series of coded pulses, it means that the velocity of light pulses just represent the velocity of information transmission. Although in the experiment of a gain aided superluminal transmission of light pulse [2,14], this pulse only transit 6 cm distortion-less, if these pulses are coded and describe a certain information, after detecting and decoding we can get this information again. Nimtz had finished the superluminal signal transmission firstly by virtue of tunnel effect of barrier [9], but that not recognized by some scientists, since they think  $v_c$  must less than  $c$ , obviously this point of view is not correct. Because of the distance of tunnel effect is very short that limit its application in actual communications, on the other hand the group velocity larger than  $c$  has a special meaning in information transmission. Recently we have given that the reason of superluminal transmission of the light pulse is due to the interaction between light and matter, and is result of reshape of pulse by medium dispersion, we also proved that by appropriate choice the parameters the superluminal pulses can transfer a long distance no distortion [19]. At present the scientists are considering that to adopt various measures to observe the superluminal transmission of light pulses, including anomalous dispersion method nearby the absorption line

[10], non-linear line method [11], exciting plasma method [12], linear gain spectral line [13], or in barrier tunnel effect to observe the superluminal light transmission [14], the fact  $v_g > c$  will be demonstrated by the more and more experiments. We think still have another measure to get higher anomalous dispersion medium.

### 2.3 Defect and Misread of Special Relativity

Now we discuss the conflict between superluminal Transmission of pulse and Special Relativity Briefly. This question we have discussed in other paper [note 1] in detail. The base of SR is following two principles:

- 1, the Principle of relativity, in all inertial systems the physical laws is equal.
- 2, the principle of invariance of light velocity, the velocity of light is independent to the movement of light source in any coordinates.

#### 2.3.1 The Causality

The soul of SR is the invariance of light velocity and the core is Lorentz transformation deduced from the invariance of light velocity. Einstein said if  $u > c$  the time inversion will happen, and it violate the causality [6]. We can prove that Einstein's conclusion is wrong. SR tells us the relationship of temporal and spatial quantities in two coordinates making a relative motion in z direction with the motive velocity u is:

$$\begin{aligned}
 x' &= x; & y' &= y, & z' &= \frac{z - ut}{\sqrt{1 - u^2 / c^2}}; \\
 t' &= \frac{t - uz / c^2}{\sqrt{1 - u^2 / c^2}};
 \end{aligned}
 \tag{5}$$

From equation (5), if  $0 < u < c$ , the denominator  $\sqrt{1 - u^2 / c^2}$  is real, and if  $u = c$ , the value of denominator is infinite, but when  $u > c$ , the value of denominator is imaginary, considering that  $i = \sqrt{-1}$ , we can rewrite the denominator as  $i\sqrt{u^2 / c^2 - 1}$ , so we have

$$t' = \frac{t - uz / c}{i\sqrt{u^2 / c^2 - 1}}
 \tag{6}$$

We can prove that, if  $|t_2 - t_1| > 0$  (in real axis), and still have  $|t'_2 - t'_1| > 0$  (in imaginary axis); it means that in such case, the events keep in order, no violation of causality happen. It is very important for information transmission. But when  $v_g < 0$ , the time advance occurs, in such case the causality is called as invert causality, but after passing such medium, these pulses keep in order too. According to the philosophic point of view, the change of quantity will induced the quality change, When velocity u varies from 0 to c then larger than c, the relationship of coordinates between two inertial systems varies from real to imaginary. Only when  $v_g < 0$ , the inversion of causality occurs, in theory we proved that is permitted.

### 2.3.2 Restriction of Lorentz Transformation

We have to say that the Lorentz transformation is only suitable in vacuum. So that the discussion of SR also only is suitable for the inertial systems in vacuum too. Because of the dragging effect of the velocity [8,10,15], in media, the Invariance of light velocity is valid only for the systems the media in which is homogenous and isotropic, also is rest respect to the observer. Especially two systems fit the invariance of light velocity simultaneously; the relative move velocity between them must be zero. The Lorentz transformation evolves into a generalized Galilean transformation. Therefore we said the SR isn't a universal theorem. It is not suitable for medium.

### 2.3.3 The Imagery Character of Special Relativity

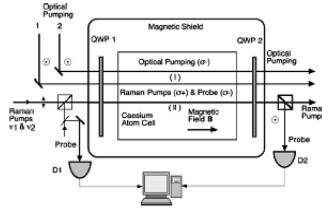
Lorentz transformation gives us the relationship of temporal and spatial quantities between two inertial systems that making relative move in vacuum. In one system, we observed only the image of object in the other system, not the object itself. The SR only gives the variation law of the coordinates of image and object with the relative moving velocity in four-dimension space, and basically not refers to the movement law of object itself. Try to get the real movement law of object from the variation of image is almost impossible. The rule contraction and clock retardation only have the apparent meaning, and they alter with the relative move velocity of two systems, and have no any actual meaning, so called father-son paradox never occur absolutely. Therefore we say that the maximum of light speed is determined by the transmission property of medium and not by the theory of SR, both of the phase velocity and group velocity can larger than  $c$  absolutely [8], the superluminal propagation of light pulses is the practical fact, and not the visual error. We also proved that the superluminal pulses could propagate a long distance distortion-less by virtue of a gain- aided medium [17].

## 3 Prospect of Superluminal Optical Communication

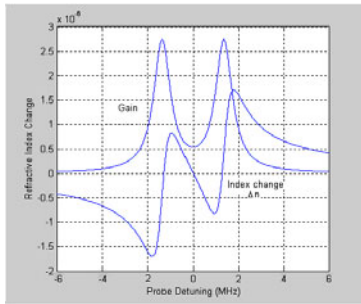
### 3.1 Advantage of the Superluminal Communication

At present, the information need to exchange increases exponentially, since the limitation of the capability of the opto-electric techniques the bite rate transmitted by each channel is about several tens G b/s, for fit the requirement of information increase day by day, a dense wavelength division multiplex (DWDM) technique has adopted, the channel number reach several tens or hundreds. In the advanced DWDM System the compensation technique of dispersion and non-linearity of fiber as well as the multiplex and de-multiplex is very complex make the system is more expensive. In fiber communication system, the velocity of signal is almost constant; its value is about  $c/1.5$ . Since the index of fiber core is  $n_1 \approx 1.5$ . If the velocity of signal is faster than  $c$ , suppose its velocity can reach as high as  $100c$ , Obviously one channel can has a several Tb/s information capacity. Single channel transmission will simplifies the compensation technique of dispersion and non-linearity, also brief the exchange and multiplex techniques, the cost of communication system will decrease. In the  $v_g < 0$  situation, since the time advance of output pulse happen, we can manufacture a "time compensator" to short the transfer time of information. From the construction of

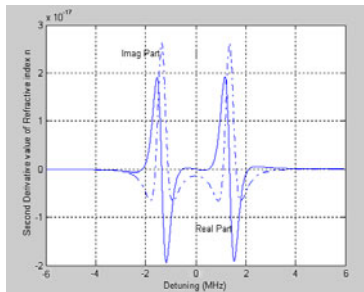
information highway, only after getting the faster signal transmission speed, we can say that we have constructed a real information highway exactly. So that to develop the superluminal communication is very attracting. We are facing the new task how to transmit a series of coded light pulses in a medium for long distance.



**Fig. 1.** Experiment setup for detecting propagation direction of light pulse when  $n_g$  is negative.[ref.3]



**Fig. 2.** Calculated value of refractive index change and gain coefficient



**Fig. 3.** Curve of second derivation of index n,— Real; - - - Imag

### 3.2 Realization of the Superluminal Communication

Since the coded pulses bring with the information propagate in medium with the group velocity, if we want  $v_g > c$ , the dispersion of medium must be negative, it means that we need the pulses propagate in anomalous region. In theory, we can choice a

suitable  $dn/d\omega$  make the  $0 < n_g < 1$ , and  $v_g = c/n_g > c$ . before the experiment [1], most of scientist obtained the larger dispersion by means of the gain or absorption spectral line, because of the frequency superposition of signal and spectral line, the distortion of the signal is serious, right now scientists still study how to get a larger negative dispersion. The measure of gain-aided superluminal propagation of light pulses is one be worth to suggest.(see fig. 1) Using two pumping Raman lasers with very near frequencies ( $\nu_{1,2} = \nu_0 \pm \Delta\nu$ ), induce two gain peaks, When the signal with a frequency  $\nu_0$  propagates in this medium, which signal also can get a gain to compensate the loss of the medium, when the density of two pumping Raman laser beam is equal, the polarizability of medium for that signal beam is [19]

$$\chi(\nu) = \frac{M}{\nu - \nu_1 + i\gamma} + \frac{M}{\nu - \nu_2 + i\gamma}. \quad (7)$$

Where  $M$  is a parameter connected with the medium character and the power of pumping beams. Based on equation (1) and (7), we can calculate the refractive index  $n$  and gain [2] (fig. 2). In WKB experiment the group velocity index measured  $n_g = -310 (\pm 5)$ , observed the group velocity is negative, that is faster than  $c$  and appear the time advance, by the proper selection of the parameters, we should get  $1 < n_g > 0$ , and  $v_g > c$ , we also demonstrated that by calculation, the superluminal light pulses can propagate long distance no distortion [18]. In fact in normal dispersion region, we can deduce the group velocity to  $8m/s$ . but the recognized light speed  $c = 2.99792458 \times 10^8$  m/s.

### 3.3 The Problems Faced by Superluminal Communications

Although the superluminal propagation of light pulses has been observed, but still has a long way far from the practical application. At first, since the medium has larger negative dispersion, the distortion of light pulses is larger and limits the long distance transmission of information, when the light pulses propagate in such medium, but this distortion can be corrected by the approach of phase conjugation [14,19], Fortunately in such case, the second derivative responding to distortion is very small in center region (fig 3). Second, since dispersion  $dn/d\omega$  depends on parameter  $M$ , therefore  $v_g$  also depends on the Intensity of pumping laser beam and Density of medium Atoms devoted to  $\chi$ . And so on, so that the time jitter is important, since the clock information also follows the signal transmission. In such case the problem of time jitter will be another research items. At last, in superluminal communication, the requirement of frequency stabilization of light source and the width of spectral line is serious; to raise the laser quality also is important, superluminal communication at future only employed in trunk or special line. But the techniques of time compensation can be applied in remote control and remote measurement in space navigation.

## 4 Summary

In this paper we only discussed the possibility of the superluminal communications technique and the problems they are facing. Here we only concentrate our discussion on

the signal transmission, of cause the techniques of signal generation and detection also need promoting to fit the requirement of the superluminal communication. For suiting the information increasing demands constantly, scientists are studying new communication techniques, such as quantum communications, but still someone doubt the quantum communications [20] Because of it need two channels. The superluminal communications rather talk by some body in present, the superluminal communication is a pure physics optical method and only one channel need, we think it is better than any faster speed communication, we hope there more persons interest in this kind of communication. Author thinks Dr. L.J. Wang give his original paper to us.

## References

1. Agrawal, G.P.: *Nonlinear Fiber Optics*, 2nd edn. Academic Press (1995)
2. Huang, Z.: *Theory and Experiments of Faster than Light Research*. Academic press, Beijing (2005) (in Chinese)
3. Wang, L.J., Kuzmich, A., Dogariu, A.: *Nature* 406, 277 (2000)
4. Zhang, Z.H., Zhang, H.A.: Cause of Superluminal Propagation of light pulse and Photon capture. In: AOM 2010, Guangzhou, China (December 2010); Zhang, Z.: Reshape and superluminal propagation of light pulse in negative dispersion medium. *Chinese Scientist* (1), 1 (2011)
5. Budker, D., Kimball, D.F., Rochester, S.M., Yashchuk, V.V.: *Phys. Rev. Lett.* 83, 1767 (1999); Xiao, M., et al.: *Phys. Rev. Lett.* 74, 666 (1995); Hau, L.V., Harris, S.E., Dutton, Z., et al.: *Nature* 397, 594 (1999); Kash, M.M., et al.: *Phys. Rev. Lett.* 82, 5229 (1999)
6. Møller, C.: *The Theory of Relativity*, 1st edn. (1955); Einstein, A.: *The Meaning of Relativity*. Princeton University Press (1953)
7. Brillouin, L.: *Wave propagation and group velocity*. Academic Press, New York (1960)
8. Zhang, Z.: Study of superluminal phenomena and its influence on Physics and Infomatics. In: 6th Photonics Conference of China, ZhongQing (August 2008)
9. Jenkins, F.A., White, H.E.: *Fundamentals of Physical Optics*, 1st edn. (1937)
10. Wang, Z.Y.: Private communication
11. Akulshin, A.M., Barreiro, S., Lezama: *Phys. Rev. Lett.* 83, 4277 (1999)
12. Basov, N.G., et al.: *Sov. Phys. JETP* 23, 16 (1966)
13. Mitchell, M.W., Chiao, R.Y.: *Am. J. Phys.* 66, 14 (1998)
14. Fisher, D.L., Tajima, T., Downer, M.C., et al.: *Phys. Rev. E* 51, 4860 (1995)
15. Steinberg, A.M., Kwiat, P.G., Chiao, R.Y.: *Phys. Rev. Lett.* 71, 708 (1993)
16. Dogariu, A., Kuzmich, A., Wang, L.J.: *Phys. Rev. A* 63, xxx (2001)
17. Nimtz, G., Heitmann, W.: Superluminal photonic tunneling and quantum electronics. *Prog. Quam. Electr.* 21(2), 81–108 (1997)
18. Zhang, Z.: Propagation of superluminal pulses in gain aided medium. In: “shenzhiguang” – Recent Anovation Theory and Practice of China. Science, technology and culture Press of China, Beijing (2009)
19. Zhang, Z.H.: *Proceeding of SPIE*, vol. 4603, pp. 105–111 (2001)
20. Wang, H.Z., et al.: Actual Nature of light and the Predication of the Rapid Development 9(3), 1–10 (2011); [Note 1]: Zhang, Z., Zhang, H.: Can the propagation of information faster than  $c$ ? *Applied Scope and Misread of Special Relativity* (manuscript, 2011)

# Design and Process Simulation on High Voltage VDMOS

Shouguo Zheng<sup>1,2</sup>, Jian Zhang<sup>2,3</sup>, Xinhua Zeng<sup>2</sup>, Zelin Hu<sup>2</sup>, and Zede Zhu<sup>2,4</sup>

<sup>1</sup> Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, China

<sup>2</sup> Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China

<sup>3</sup> Wuxi Intelligent Agricultural Development CO. LTD.

Chinese Academy of Sciences, Wuxi, China

<sup>4</sup> University of Science and Technology of China, Hefei, China

zhshg1985@163.com, {jzhang, xhzeng}@iim.ac.cn,  
{hzlin, zhuzede}@ustc.edu

**Abstract.** A VDMOS cell structure based on semiconductor physics and relevant states were designed. The process parameters of the high-voltage VDMOS have been calculated and adjusted. Using the optimal device' dimensions, the doping concentration was analyzed systematically. Additionally, a simulating model of VDMOS was compiled based on SENTAURUS PROCESS. According to this model simulation, an optimized device parameters' design was further determined. Finally, the device drain-source breakdown voltage has satisfied a standard value of 1200V. Whereas its threshold voltage is 3.9V, which is in the prerequisite range of 3~ 5V.

**Keywords:** VDMOS, Breakdown voltage, ON-resistance, Sentaurus tcad.

## Introduction

With the rapid development of power electronics technology, the role of high voltage VDMOS (vertical double diffused metal oxide semiconductor) device is becoming an issue of increasing interests. To our knowledge, the design and manufacture of VDMOS power device are in the elementary stage. The research on high-voltage VDMOS has not been mature enough. The purpose of this paper is to design and simulate VDMOS power device model, and to find the best structure through Computer simulation and analysis.

## 1 Structure Design of VDMOS

The VDMOS structure is as shown in Fig.1. In a VDMOS structure, an n<sup>-</sup>type epitaxial layer is formed on an upper surface of an n<sup>+</sup>-type substrate and p-type diffusion regions are selectively formed on its upper surface, while n-type diffusion regions are further formed on upper surfaces. A gate electrode wrapped in an oxide film is covered on the upper surface of the n<sup>-</sup>type epitaxial layer and the p-type diffusion regions. The gate is held between the n<sup>-</sup>type epitaxial layer and the n<sup>+</sup>-type diffusion regions.



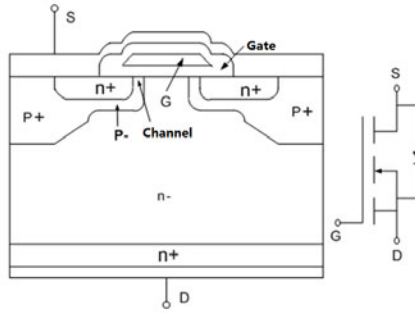


Fig. 1. VDMOS cell structure

In order to improve the ruggedness of the device, the ability to withstand an avalanche current during an unclamped inductive load switching event must be improved. Simultaneously, the turn-on of the parasitic drain-body-source npn bipolar junction transistor (BJT) must be suppressed.

### 1.1 Design of N- Drift Region

The voltage handling capability of VDMOS structure is given by the breakdown voltage ( $V_B$ ) of the P-body/N–junction, that is strongly dependent on the thickness ( $W_{epi}$ ) and the doping of the epitaxial layer ( $N_{epi}$ ).

$$V_B = E_c W_{epi} - \frac{q N_{epi} W_{epi}^2}{2 \epsilon_0 \epsilon_s} \quad W_{epi} < W_M$$

To sustain high voltage, the VDMOS requires a low doping and a thick epitaxial layer. However, both low doping and thickening result in a large ON-resistance in the on-state.

$$R_{on} = R_0 A = \rho \frac{W_{epi}}{A} A = \frac{W_{epi}}{q \mu_n N_{epi}}$$

Advances in the process technology have improved transistor packing density and, consequently, transistor specific on-resistance ( $R_{on}$ ), which is the product of the on-resistance and the active area of the device. However, improvements in specific on-resistance have been limited by material and breakdown voltage, which required a relative thick and low doping epitaxial layer. The theoretical limit, called "silicon limit", for the " $R_{on}$  versus  $V_B$ " trade-off of conventional VDMOS is given by

$$R_{on} = 8.3 \times 10^{-9} \cdot V_B^{2.5}$$

According to the simulation of VDMOS breakdown voltage in different drift region doping concentration, the results is as shown in Table 1:

**Table 1.** The relationship between concentration of N- and  $V_B$ 

concentration of N- ( $cm^{-3}$ )	$8.0 \times 10^{13}$	$8.7 \times 10^{13}$	$9.5 \times 10^{13}$	$1.0 \times 10^{14}$
$V_B$ (V)	1520	1450	1220	1180

The design goal in this paper is to 1200V, taking into account the terminal and process implementation issues, a certain margin must be left, so the design refer to 1500V. As a result:

$$N_{epi} = 8.7 \times 10^{13} \text{ cm}^{-3}$$

$$W_{epi} = 112 \mu\text{m}$$

## 1.2 Design of Gate Oxide

VDMOS gate oxide thickness is one of the crucial parameter, which directly determines the threshold voltage  $V_{th}$ , and it also determines other two important parameters: gate charge  $Q_g$  and gate capacitance  $C_g$ . Through the simulation of relationship between Gate oxide thickness and threshold voltage, the results is as shown in Table 2:

**Table 2.** The dependence of Gate oxide on  $V_{th}$ 

Gate oxide ( $\text{\AA}$ )	300	400	500	600
$V_{th}$ (V)	3.5	3.9	4.3	4.5

## 1.3 Design of N+ Source

VDMOS source using high doses of phosphorus (P) or arsenic (As) injection, so it can make a connection with the source region of the Schottky barrier between? the metal below. Therefore, it can reduce the ohmic contact resistance. It can be observed from the simulation software to 1200V VDMOS device that the impact of the junction depth of source region to the threshold voltage  $V_{th}$  is shown in Table 3:

**Table 3.** The dependence of Junction depth on  $V_{th}$ 

Junction depth ( $\mu\text{m}$ )	0.3	0.4	0.5
$V_{th}$ (V)	3.84	3.90	3.95

## 1.4 Design of the P-Body

B element was first injected into the the trap. Upon heating, it is then diffused through the  $n^-$  region, which gives rise to the P-body after annealing. The P-body have two key parameters: the junction depth and implantation dose. The implantation dose

greatly influences the surface of the channel impurity concentration, thus directly affects the  $V_{th}$ . Theoretically, the dependence of  $V_{th}$  on the channel impurity concentration is as shown:

$$V_{th} = -(\phi_{MS} + \frac{Q_{ox}}{C_{ox}}) + \frac{\sqrt{4qN_{ch}\epsilon_{si}\epsilon_0\phi_F}}{C_{ox}} + \frac{2k_0T}{q} \ln \frac{N_{ch}}{n_i}$$

And the result observed from the simulation software is shown in Table 4:

**Table 4.** The relationship between P-body dose and  $V_{th}$

P-body dose (cm <sup>-3</sup> )	5*10 <sup>16</sup>	1*10 <sup>17</sup>	2*10 <sup>17</sup>	3*10 <sup>17</sup>
$V_{th}$ (V)	3.5	3.9	4.2	4.8

### 1.5 Results of Structural Design

According to previous discussions, the final device structure parameters determined as follows:

**Table 5.** Device structure parameters

Project	Parameters
concentration of N- drift region	8.7*10 <sup>13</sup> cm <sup>-3</sup>
Channel length	1.8um
Source length	3um
Drain length	10um
Drift region length	112um
concentration of P- channel	1*10 <sup>17</sup> cm <sup>-3</sup>
concentration of P+ region	1*10 <sup>19</sup> cm <sup>-3</sup>
concentration of Source N+	1*10 <sup>20</sup> cm <sup>-3</sup>
concentration of drain N+	5*10 <sup>19</sup> cm <sup>-3</sup>

## 2 VDMOS Fabrication Process

The reference VDMOS was based on a 0.5 mm process. The starting wafer is a <100> oriented, n<sup>+</sup>-type wafer with a nominal arsenic-doping concentration of 5\*10<sup>19</sup>cm<sup>-3</sup>. At the beginning of the fabrication process, a phosphorus-doped n<sup>-</sup> layer with the concentration of 8.7\*10<sup>13</sup>cm<sup>3</sup> was epitaxially grown on the silicon wafer. Subsequently, a thick layer of oxidation area is formed by active lithography and oxide etching. field oxidation is carried out to form a thick layer of oxide followed by active lithography and oxide etching. After gate oxidation, polysilicon deposition, doping and annealing, gate lithography and polyetching, a hexagon mesh gate pattern is formed. The p-body is formed by a self-aligned implantation of boron during

annealing, while the  $n^+$ -source is engendered by a self-aligned implantation of arsenic during annealing. The lateral diffusion difference of the p-body and  $n^+$ -source forms a controlled channel length along the Si-surface. The implanting dose is based on diffusion trials and extensive process and device simulation. A masked high dose boron implantation is carried out to form a  $p^+$ -region in the p-body to enhance body contact. Subsequently, a thick intermediate oxide layer of tetraethyl orthosilicate is deposited after the formation of the contact window using contact lithography and oxide etching. Finally, the chip surface is deposited a metal layer to form the butting source/body electrodes for the VDMOS. By distributing metal contacts on the polysilicon gate around the edge of the chip, the device's gate resistance is minimized.

The fabrication process of VDMOS is shown in Fig.2. After the simulation using Sentaurus Technology Computer-Aided Design software (STCD), which refers to the use of computer simulations to develop and optimize semiconductor processing technologies and devices, the final simulation result is shown in Fig.3.

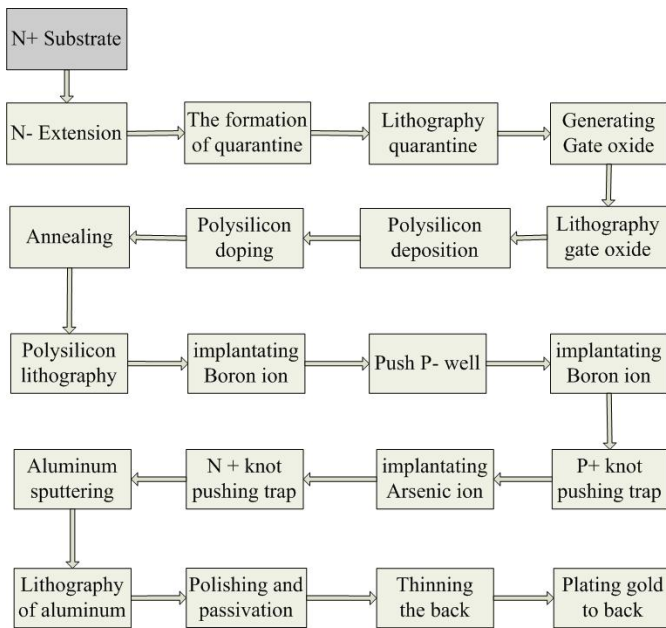


Fig. 2. VDMOS fabrication process

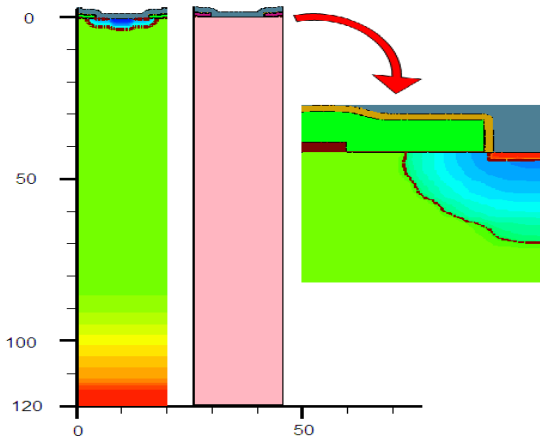


Fig. 3. Simulation results

### 3 Device Analysis

#### 3.1 Important Parameters Simulation to VDMOS

The VDMOS can be regarded as both a power device and a MOS pipe device. Accordingly, three important parameters is involved in this simulation. As a power device, breakdown voltage and ON-resistance are the most important two factors. Concerning a MOS pipe device, the primary parameter is the gate turn-on voltage. The following simulation study focuses on these three key characteristics.

The breakdown characteristic is shown in Fig.5, the breakdown voltage is 1450V, which attains the desired standard.

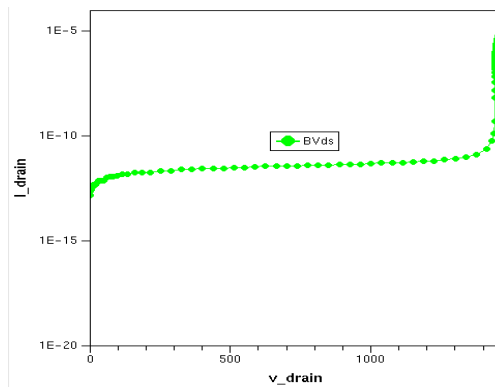


Fig. 4. Breakdown characteristic

Device on-resistance is shown in Fig.5, gate voltage is fixed in 5V, and the drain voltage is swept. It can be seen from Fig.5: the relationship between on-resistance and

drain voltage is logarithmic. Transfer characteristic of the device is shown in Fig.6, it can be seen from the figure that the threshold voltage is 3.9V. When the gate voltage  $V_{gs}$  reach 4.8V, the drain current  $I_{drain}$  saturated, it is the different place between VDMOS pipe and common pipe in transfer characteristics.

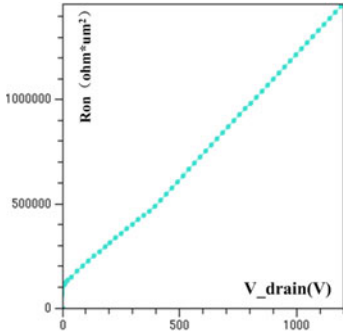


Fig. 5. On-resistance

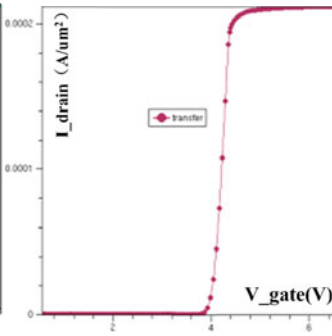


Fig. 6. Transfer characteristic

### 3.2 Quasi-saturation Effect

When VDMOS devices at high gate-source voltage or high drain-source voltage, drain current is not saturated, and this is the VDMOS transistor quasi-saturation effects. It will limit the applications of VDMOS transistors in high power. This quasi-saturation effect of the VDMOS device designed in this paper is simulated as shown in Fig7.

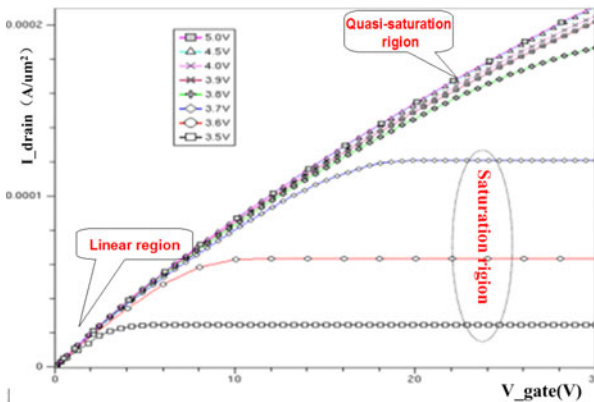


Fig. 7. Quasi-saturation effect

VDMOS quasi-saturation limits the maximum output current, It results from the carrier velocity saturation in the neck area, the output current saturation is proportional to the distance between the P-body area trap. Quasi-saturation effect

caused the following consequences: 1. transistor output current has an intrinsic limit; 2. When power VDMOS transistors used in the quasi-saturated zone, its transconductance tends to zero.

### 3.3 Temperature Characteristics

Since VDMOS is high-voltage power device, large temperature change will occur frequently during the work. Consequently, the study of the temperature characteristics of VDMOS is essential. The work focuses on the impact of temperature to the ON-resistance and threshold voltage.

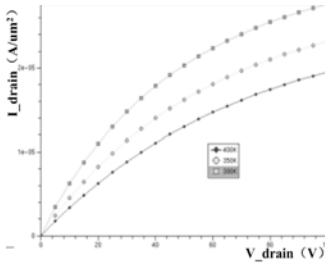


Fig. 8. Influence of temp to on-resistance

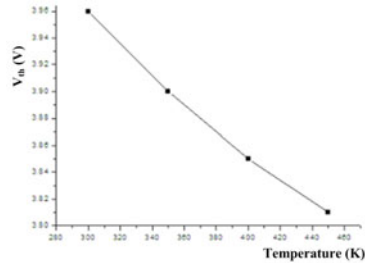


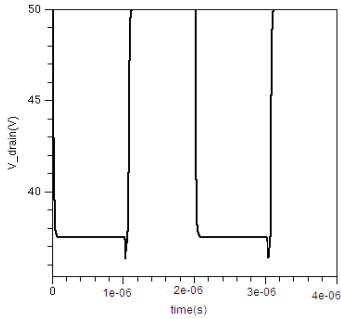
Fig. 9. Influence of temp to V<sub>th</sub>

It can be seen from the fig.8, as the temperature increase, the drain current has a downward trend, this is because the resistance becomes larger with the increasing of temperature. From Fig.9, it can be seen that at a temperature of 300K, 350K and 400K, the threshold voltage V<sub>th</sub> is decreased as the temperature increase, due to the structure of the substrate concentration is not very high, so the impact is not obvious when the temperature ranging is little.

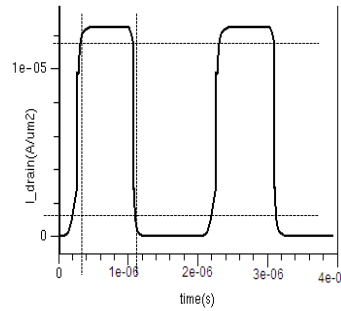
### 3.4 Dynamic Characteristic

VDMOS transistor switching speed is much faster than the same size of the bipolar transistor, because it use the majority carrier conduction, and does not require storage in the on-off time. So the switching waveform is influenced by the internal electrode charging and the discharging of the capacitor. The switching characteristics of VDMOS device designed in this paper are shown as following Figures.

It can be seen from Fig.10 that the rising edge of the output voltage V<sub>drain</sub> has a great pulse, it is due to discharge capacitor C<sub>ds</sub>. VDMOS turn-on-time is defined as the time that the current value rise from 0 up to the current conduction when the 90% of current value I<sub>max</sub>, turn-off-time is defined as the time that the current from the conduction current I<sub>max</sub> decreased to 10% of I<sub>max</sub>. It can be seen from the figure.9 that turn-on-time is about 0.32s, turn-off-time is about 0.12s.



**Fig. 10.** Drain voltage versus time



**Fig. 11.** Drain current versus time

## 4 Conclusion

A 1200V high voltage VDMOS structure has been designed, which is based on the analysis of the essential properties and working principle of VDMOS power device. The parameters of the designed structure were analyzed and simulated, and the work of this paper will provide a theoretical basis to the production of high voltage VDMOS.

**Acknowledgments.** The authors thank the National Natural Science Foundation of China (Grant No. 21107114) and National 863 plans projects (SQ2010AA1000764003), for the financial and technical support.

## References

1. Young, S.K., Lee, J.S.: A Novel Trench IGBT with a Deep P+ Layer Beneath the Trench Emitter. *IEEE Electron Device Letters*, 82–84 (2009)
2. Xu, H.P.E., Trescases, O.P., Sun, I.-S.M., Lee, D.: Design of a rugged 60 V VDMOS transistor. *IET Circuits Devices Syst.*, 327–331 (2007)
3. Philips Semiconductor Applications AN10273\_1. Power MOSFET single-shot and repetitive avalanche ruggedness rating. Philips Semiconductors (2003)
4. Paredes, J., Hidalgo, S., Berta, F.: A Steady-State VDMOS Transistor Model. *IEEE Trans. Electron Devices*, 712–714 (1992)
5. Johnson, C.M.: Current state-of-the-art and future prospects for power semiconductor devices in power transmission and distribution applications. *Int. J. Electron.*, 667–693 (2003)
6. He, J., Chen, X.: Increasing breakdown voltage of LDMOST using buried layer. *Chinese Journal of Semiconductors*, 753–758 (1999)



# Parameters Estimation of Stratospheric Airship and Model of Inflating and Deflating Air

Shaoxing Hu<sup>1</sup>, Erbing Wu<sup>1</sup>, and Aiwu Zhang<sup>2</sup>

<sup>1</sup> School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China

<sup>2</sup> Key Laboratory Of 3D Information Acquisition And Application, Capital Normal University, Beijing 100037, China

husx@buaa.edu.cn, tjwubing@yahoo.com.cn, Zhangaw98@163.com

**Abstract.** Stratospheric airship is a complicated system. Parameters estimation and flight control are two important sides during the design of stratospheric airship. According to components of airship system, partial parameters were estimated from solar battery, propulsion system, buoyancy gas and so on, and a conceptual model of stratospheric airship was established. In accordance with the model established, based on the consistency of pressure difference between internal and external of stratospheric airship, a model of inflating and deflating air in the period of ascent and descent was established and simulated. The result indicates that the temperature of ballonet, the height and speed of flying have great influence to the velocity of inflating and deflating air.

**Keywords:** Stratospheric airship, Parameters estimation, conceptual model, pressure difference, model of inflating and deflating air.

## 1 Introduction

In recent years, with the development of space science and technology, stratospheric space resource has been developed and utilized by human. And with the growth of human's needs for space observation, atmospheric measurement, military surveillance and so on, a concept of stratospheric platform was put forward. Stratospheric platform is a kind of aerostat that can reside in stratosphere for a long time. According to different loads carried, stratospheric platform can execute different tasks. Many countries attach importance to research on stratospheric platform. America, Japan, Korea and some Western European countries have carried out feasibility studies and experiments and made great progress. At present, China does little research on stratospheric platform.

As a kind of major stratospheric platform, stratospheric airship composes of hull, keel, ballonets, empennages, rudders, solar battery, propeller, payload and so on [1,2]. Different from other aircrafts, most lift force of airship derives from buoyancy generated by gas that lighter than air in ballonets, and small part provided by propeller. Weight and volume are keys to the design of airship, meanwhile, energy and power needs must be taken into account. Besides, environment parameters such

as pressure, temperature and density of atmosphere change with time and height. In the period of ascent, hovering and descent, we need make adjustment duly to ensure that airship flies smoothly.

Through the estimation of solar battery, propulsion system and buoyancy gas, a model of airship was established. Then based on the model, and assuming that the airship flies in a constant speed, a model of inflating and deflating air in the period of ascent and descent was established.

## 2 Estimation of Population Parameters

In general, airships can be divided into non-rigid airships, rigid airships and semi-rigid airships [3].

Non-rigid airship has multiple air ballonets and helium ballonets. Its shape is sustained by the pressure of internal gas. Rigid airship has a full keel structure that sustains the shape of airship. It has stronger capability of resisting wind pressure. Semi-rigid airship is between non-rigid airship and rigid airship. It is produced by installing keel at the main parts that bear force in non-rigid airship.

### Structure Shape of Airship

Before establishing the model of airship, we assume that volume is  $V_{\text{艇}}$ ; surface area is  $S_{\text{艇}}$ ; payload is  $m_p$ .

Generally, the hull is designed into the shape of water-drop, which benefits reducing air resistance during flight. It is approximately composed of two semi-elliptical [4]. Short axis of the two ellipses are the same, but long axis not. Assuming that half of short axis is  $b$ ; half of long axis are respectively  $a_1$  and  $a_2$ ; slenderness ratio of airship is 3.5, then:

$$V_{\text{艇}} = \frac{2}{3}\pi(a_1 + a_2)b^2 \quad (1)$$

$$S_{\text{艇}} = \frac{2}{3}\pi b(2a_1 + 2a_2 + 2b) \quad (2)$$

$$f = \frac{a_1 + a_2}{2b} = 3.5 \quad (3)$$

Buoyancy of airship is:

$$F_{\text{浮}} = \rho_a V_{\text{艇}} g \quad (4)$$

Where  $\rho_a$  is the density of atmosphere.

### Mass Composition of Airship

As mentioned before, an airship consists of hull, keel, ballonets, propeller and so on, so the total mass is:

$$m_{\text{艇}} = m_{mp} + m_{qn} + m_{He} + m_{Air} + m_{tf} + m_{rf} + m_{xf} + m_{pro} + m_s + m_p \quad (5)$$

Where:

Mass of skin  $m_{mp} = \rho_{mp} S_{mp}$ ,  $\rho_{mp}$  is surface density of skin,  $S_{mp}$  is area of skin;

Mass of ballonnet  $m_{qn} = \rho_{qn} S_{qn}$ ,  $\rho_{qn}$  is surface density of ballonnet,  $S_{qn}$  is area of ballonnet;

Mass of helium  $m_{He} = p_{\text{内}} M_{He} V_{He} / \bar{R} T$ ,  $p_{\text{内}}$  is internal pressure of airship,  $M_{He}$  is relative molecular mass of helium,  $\bar{R}$  is universal gas constant,  $T$  is internal temperature of airship,  $V_{He}$  is volume of helium;

Mass of air  $m_{Air} = p_{\text{内}} M_{Air} V_{Air} / \bar{R} T$ ,  $M_{Air}$  is relative molecular mass of air,  $V_{Air}$  is volume of air;

Mass of solar battery  $m_{rf} = \rho_{rf} S_{rf}$ ,  $\rho_{rf}$  is surface density of solar battery,  $S_{rf}$  is area of solar battery;

$m_{rf}$  is mass of fuel cell;  $m_{xf}$  is mass of storage battery;  $m_{pro}$  is mass of propulsion system, including electromotor and propeller;  $m_s$  is other rest mass, including keel, empennages and so on;  $m_p$  is mass of payload.

## Propulsion System

The main lift force of airship is its own buoyancy, but the thrust generated by propulsion system is also important in overcoming air resistance. In the process of flight, the power provided by propeller is:

$$P = 0.5 \rho_a v^3 S C_d \quad (6)$$

Where  $\rho_a$  is density of atmosphere around airship;  $v$  is flight speed of airship;  $S = V_{\text{艇}}^{2/3}$  is reference area of airship;  $C_d$  is drag coefficient of airship, which relates to slenderness ratio and Length Reynolds Number of airship.

Power of electromotor is:

$$P_m = P / \eta \quad (7)$$

Where  $\eta = \eta_p \eta_t \eta_m$  is transmission efficiency from electromotor to propeller.  $\eta_p$  is propulsive efficiency of propeller;  $\eta_t$  is efficiency of transmission mechanism;  $\eta_m$  is efficiency of electromotor.

## Solar Battery

All electrical equipments on airship rely on solar battery, fuel cell and storage battery to provide energy. During the day, solar battery converts solar energy into electrical

energy, meanwhile, water is electrolyzed into hydrogen and oxygen and then stored. At night, under the action of fuel cell, chemical reaction between hydrogen and oxygen takes place to generate electrical energy and water. Storage battery is used in case of emergencies such as insufficient electrical energy.

With regard to horizontal plane perpendicular to sunray, if not take atmospheric absorption and scattering into account, direct radiation that the plane receives from the sun is [5, 6]:

$$E_h = E_{sc} \cdot (r_0 / r)^2 \quad (8)$$

Where  $E_{sc}$  is solar constant, generally  $E_{sc} = 1367W / m^2$  ;  $(r_0 / r)^2$  is Sun-Earth distance correction factor:

$$(r_0 / r)^2 = 1.000423 + 0.032359 \sin \theta + 0.000086 \sin 2\theta - \\ 0.008349 \cos \theta + 0.000115 \cos 2\theta \quad (9)$$

where  $\theta = 2\pi(n-1)/365$  ,  $\theta = 0^\circ$  on January 1st every year; n is the nth day in a year.

To meet the energy need of airship, we choose winter solstice (December 22<sup>nd</sup>,  $n = 356$  ), in which day the sunshine duration is the shortest, to calculate,  $E_h = 1349W / m^2$  .

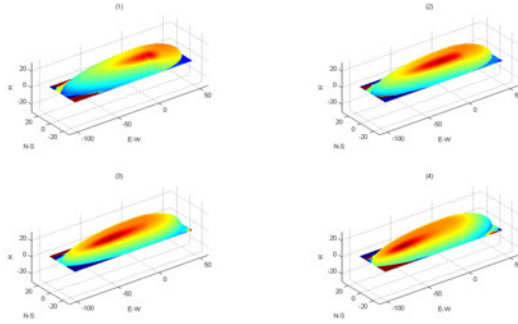
Direct radiation from the sun received by the plane in either direction is:

$$E_t = E_h [\sin \delta (\sin \phi \cos \beta - \cos \phi \sin \beta \cos \gamma) + \\ \cos \delta \cos \omega (\cos \phi \cos \beta + \sin \phi \sin \beta \cos \gamma) + \\ \cos \delta \sin \beta \sin \gamma \sin \omega] \quad (10)$$

Where  $\phi$  is the geographic latitude;  $\beta$  is dip angle of the plane ( $0^\circ$  for horizontal plane; positive for plane facing to the equator; negative for plane backing on the equator);  $\gamma$  is azimuth angle of the plane ( $0^\circ$  for plane facing south; negative for plane facing east; positive for plane facing west);  $\omega$  is hour angle of the sun, which changes in a cycle of day and night,  $\omega = (t-12) \times \pi / 12$  ( $t$  is local time from 0 to 24);  $\delta$  is declination angle of the sun (angle between terrestrial equator plane and the ligature that connects centers of earth and sun), which changes in a cycle of year.

$$\delta = 0.006918 - 0.399912 \cos \theta + 0.070257 \sin \theta - \\ 0.006758 \cos 2\theta + 0.000907 \sin 2\theta - \\ 0.002697 \cos 3\theta + 0.001480 \sin 3\theta \quad (11)$$

Because the direction of wind in stratosphere is basically east-west, the attitude of airship during hovering is east-west direction. From sunrise to sunset, sun radiation changes constantly for every part of airship surface as shown in figure 1:



**Fig. 1.** Solar radiation

With regard to airship with ellipsoid surface, according to traditional definition of coordinate system (center of buoyancy is the origin; axis  $x$  directs head of airship; axis  $z$  directs top of airship; axis  $y$  defined by Right-hand Rule), unit normal vector at either point on surface of airship is:

$$e = (\cos \alpha', \cos \beta', \cos \gamma') \quad (12)$$

Then dip angle  $\beta$  and azimuth  $\gamma$  of tangent plane at the point can be acquired by formulas as follows:

$$\begin{cases} (\cos \beta)^2 = \frac{(\cos \gamma')^2}{(\cos \beta')^2 + (\cos \gamma')^2} \\ (\cos \gamma)^2 = \frac{(\cos \beta')^2}{(\cos \alpha')^2 + (\cos \beta')^2} \end{cases} \quad (13)$$

Substitute  $\beta$  and  $\gamma$  into formula (10), and integrate on surface of airship  $\iint E_t dS dt$  ( $S$  is the coverage area of solar battery;  $t$  is time from sunrise to sunset). In this way, we can calculate the area of solar battery according to energy demand.

## Helium and Air

The main lift force of airship derives from buoyancy generated by helium. Volume and weight of airship determine the quantity of helium needed. Internal pressure of airship determines the ratio of helium and air. So, when estimate the mass of helium, we should consider the factors of volume, weight and pressure comprehensively.

In order to ensure that airship doesn't deform with wind disturbance, internal pressure of airship must exceed external pressure, that is, maintains pressure difference. The minimum pressure difference can't be less than 200Pa, and the maximum can't exceed 1000Pa [7].

In the process of ascent, to maintain pressure difference, air in ballonets needs to be deflated constantly. So, volumes of helium and air change constantly. Reaching the maximum height, the volume of helium is equal to that of the hull more or less. Considering the change of temperature difference and volume occupied by other

control systems, there is usually about 5% of the volume of hull reserved to adapt to volume fluctuation of helium caused by temperature change and accommodate other objects. Assuming volume occupied by other subsystems is  $\nabla$ , at the height of 20km, the volume of helium is:

$$V_{20000\text{氦}} = 95\%V_{\text{艇}} - \nabla \tag{14}$$

Density of helium can be acquired by the equation of gas state.

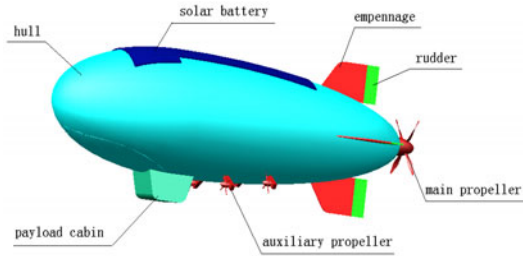
**Model of Airship**

At the height of hovering, to realize that airship maintains the height, gravity and buoyancy of airship are equal, that is:

$$\rho_a V_{\text{艇}} g = (m_{mp} + m_{qn} + m_{He} + m_{Air} + m_{lf} + m_{rf} + m_{xf} + m_{pro} + m_s + m_p) g \tag{15}$$

$$V_{\text{艇}} = V_{\text{氦}} + V_{\text{空}} + \nabla \tag{16}$$

Based on the method of parameters estimation about the main parts, a model of airship is acquired preliminarily, as shown in figure 2:



**Fig. 2.** Conceptual model of airship

Design parameters of airship are as shown in table 1:

**Table 1.** Design parameters of airship

<b>Parameters of airship</b>	<b>Value</b>
Maximum diameter of airship( $m$ )	41
Length of airship( $m$ )	143.5
Surface area of airship( $m^2$ )	14083
Volume of airship( $m^3$ )	126304
Mass of airship( $kg$ )	8574
Mass of helium( $kg$ )	1559

Based on the volume of airship and mass of helium, a model of inflating and deflating air during ascent and descent is established.

### 3 Model of Inflating and Deflating Air

As mentioned before, environmental parameters change with time and height. If the internal pressure of airship retains constant, pressure difference between internal and external of airship will change a lot. To ensure the safety of airship, we need adjust the internal pressure with the change of external pressure.

According to the equation of gas state  $p = \rho RT$  [8], there are two methods to achieve the pressure regulation: 1) adjusting the temperature; 2) adjusting the density of gas. For gas with constant mass, we can adjust its volume. At present, the second method is more used.

#### Volume Change of Helium

Throughout the process of flight, assuming that helium doesn't leak, we can adjust the internal pressure by adjusting the volume of helium [9]. In the period of ascent, we assume that the pressure difference is 500Pa constant. Because the atmospheric pressure changes with height, the volume of helium will change as well:

$$V_{\text{氦}} = \frac{m_{\text{氦}} \bar{R} T}{(p_{\text{外}} + \Delta p) M_{\text{He}}} \tag{17}$$

Where  $m_{\text{氦}}$ ,  $\bar{R}$ ,  $m_{\text{He}}$  are constants;  $T$  and  $p_{\text{外}}$  relate with the height  $H$ . Substituting the value of every parameter into the formula, we can obtain the relationship between the volume of helium and the height:

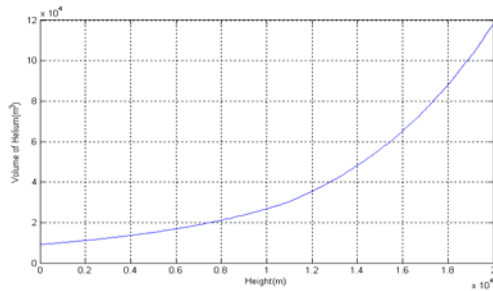


Fig. 3. Volume change of helium with height

As shown in figure 3, the volume of helium increases constantly with the height increases. The volume of helium reaches to the maximum more or less when airship flies to the height of 20km.

The volume model of helium above is established under the circumstance that inside and outside of airship have the same temperature. In fact, because the airship is

bulky, influenced by solar radiation, there is temperature difference between inside and outside of airship. If the temperature difference is 4K, the pressure difference will increase 1414Pa on the ground, which exceeds the safety range of airship. At this moment, to insure the safety of airship, we need deflate some air to decrease internal pressure. So, volume of helium will increase somewhat.

During the flight, the external forces that airship receives include buoyancy, gravity, air resistance, fluid inertia and thrust [10]. At a certain height, temperature change leads to pressure change. If we adjust the internal pressure by deflating air, the weight of airship will decrease. Because other external forces also change with height, it is necessary to adjust the power system to ensure the airship can rise or fall in a constant speed.

### Model of Inflating and Deflating Air

According to figure 3, in the case of constant volume of airship, the volume of air ballonets decrease with the height increases.

Assuming that the airship rises in a constant speed of  $v$  and the velocity of deflating air is  $q_1(t)$ , the height increases linearly with time:

$$H = vt \tag{18}$$

Volume of air is:

$$V_{\text{air}} = V_{\text{air}0} - \int_0^t q_1(x) dx \tag{19}$$

Where  $V_{\text{air}0}$  is the volume of air on the ground.

Substitute formulas (17) and (19) into formula (16):

$$V_{\text{air}} - (V_{\text{air}0} - \int_0^t q_1(x) dx) - \nabla = \frac{m_{\text{air}} \bar{RT}}{(p_{\text{ext}} + \Delta p) m_{\text{He}}} \tag{20}$$

Calculating time derivative on both sides of the equation, we can obtain the relationship between  $q_1(t)$  and  $t$ , and then get the relationship between  $q_1(t)$  and  $H$ , as shown in figure 4:

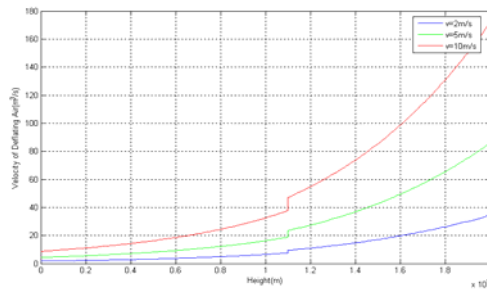
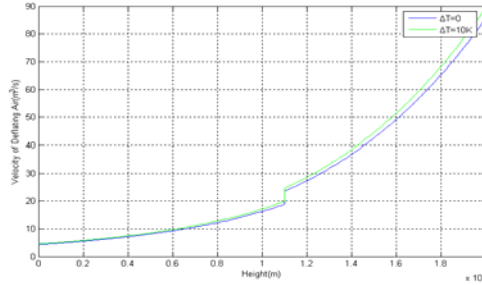


Fig. 4. Velocity of deflating air



According to figure 4, for different speeds of ascent, the velocities of deflating air are different. The faster airship ascends, the faster velocity of deflating air is. The reason is that the faster airship ascends, the faster surrounding atmospheric pressure decreases, the faster volume of helium increases, and it is needed to deflate air faster. For a constant speed, the velocity of deflating air increases constantly with the height increases. In addition, when  $H$  reaches to 11km, a rapid transition occurs for the velocity of deflating air. It is because the atmospheric parameters are not smoothly continuous there.

Temperature affects the change of volume of helium. So, it will also affect velocity of deflating air. With the temperature rises and volume of helium increases, velocity of deflating air increases, as shown in figure 5:



**Fig. 5.** Temperature influence to velocity of deflating air

During the period of descent, assuming that the airship descends in a constant speed of  $v$  and the velocity of inflating air is  $q_2(t)$ :

$$H = 20000 - vt \quad (21)$$

Volume of air is:

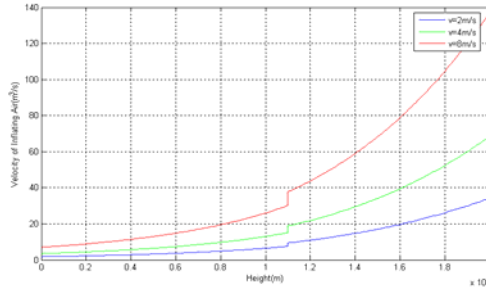
$$V_{\text{空}} = V_{\text{空}20000} + \int_0^t q_2(x) dx \quad (22)$$

Where  $V_{\text{空}20000}$  is the volume of air at the height of 20km.

Substitute formulas (17) and (22) into formula (16):

$$V_{\text{艇}} - (V_{\text{空}20000} + \int_0^t q_2(x) dx) - \nabla = \frac{m_{\text{氦}} \bar{R} T}{(p_{\text{外}} + \Delta p) m_{\text{He}}} \quad (23)$$

Calculating time derivative on both sides of the equation, we can finally get the relationship between  $q_2(t)$  and  $H$ , as shown in figure 6:



**Fig. 6.** Velocity of inflating air

From figure 4 and figure 6, we know that the velocity of inflating air during ascent and velocity of deflating air during descent change the same with height. It is because the pressure determines the volume of gas in airship. Since the atmospheric pressure is consistent at the same height, velocities of inflating and deflating air are equal for the same flight speed.

## 4 Conclusion

Stratospheric airship is different from other aircrafts. In the process of flight, the mass of airship changes constantly. During the design of airship, we need estimate the weight of every subsystem of airship and determine the size of airship.

In this paper, we got relevant parameters of airship preliminarily and established the model of inflating and deflating air. We can see from the model that in the case of helium with constant mass, the factors that affect velocities of inflating and deflating air not only include the height and flight speed, but also temperature and pressure difference. Actual atmospheric environment is very complex, and it is necessary to control pressure, temperature and dynamic comprehensively to ensure safe flight of airship. This is the future research need to do.

**Acknowledgments.** This work was supported by the National Natural Science Foundation(40601081,41071255),National Defense Foundation Advanced Research (SDDX2008117) and National Education Department Doctor Fund (20070006031).

## References

1. Ilcev, S.D., Singh, A.: Development of Stratospheric Communication Platforms (SCP) for Rural Applications. In: 7th AFRICON Conference in Africa, pp. 233–238 (2004)
2. Xia, Z.-X., Xu, H.-M.: Research on Performance and Technology of Stratospheric Airship. Nanjing University of Aeronautics and Astronautics (2006)
3. Gao, M., Wu, Q.-X., Jiang, C.-S., et al.: Stratospheric Large Airship Conceptual Design and Modeling. *Aero Weaponry* (5), 10–14 (2007)

4. Mueller, J.B., Paluszek, M.A., Zhao, Y.-Y.: Development of an Aerodynamic Model and Control Law Design for a High Altitude Airship. In: AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit, September 20-23 (2004)
5. Wang, H.-F., Song, B.-F., Su, J.-M., et al.: Computation Research on The Solar Radiation on The Thin Film Solar Cell for High Altitude Airships. *Acta Energiae Solaris Sinic* 27(8), 819–823 (2006)
6. Shi, J.-H., Zhu, X.-J., Sui, S.: Modeling and Analysis of Solar Energy Collected by Stratosphere Airship's Surface. *Computerised Simulation* 24(8), 64–67 (2007)
7. Zheng, W., Wang, W.-J., Yao, W., et al.: Steady State Modeling and Characteristics Analysis of Vertical Motion for Stratospheric Airships. *Journal of System Simulation* 20(24), 6830–6833 (2008)
8. Li, F.: Introduction of air and air dynamics, pp. 11–13. Northwestern Polytechnical University Press, Xi'an (2007)
9. Gomes, S.B.V., Ramos Jr, J.G.: Airship Dynamic Modeling for Autonomous Operation. In: International Conference on Robotics & Automation, Leuven, Belgium, p. 3463 (May 1998)
10. Jin, O., Xi, Y.-G.: Modeling of Unmanned Airship and Research of Control Methods. Ph.D. thesis, Shanghai Jiao Tong University (2003)

# A New Method for Improving Life Time in Wireless Sensor Network by Using Genetic Algorithm

Shahram Babaie and Arash Khadem

Department of computer, Islamic Azad University, Tabriz Branch, Tabriz, Iran  
{Hw.tab.au, Akhadem65}@gmail.com

**Abstract.** Routing protocols in wireless sensor networks (WSN) have been substantially investigated by researchers. Most survey subject have focused on reviewing the different routing schemes that have been proposed for WSN and classifying them based on the network's type and protocol's operation also a fundamental issue in the design of Wireless Sensor Networks is to maximize their lifetimes particularly when they have a limited and non- substitute energy supply .To improve the network lifetime, controlling of power management and energy-efficient communication techniques at all sensors are necessary. In this paper, we propose a new method for the routing issue by using genetic algorithm for WSNs. Our purpose is to maximize the network lifetime by improving efficient routing techniques. Simulation results show that when we use good coefficient of our parameters, the life time of wireless sensor network is increased.

**Keywords:** Wireless Sensor network, Routing; Lifetime, Low power, Genetic Algorithm.

## 1 Introduction

The wireless sensor network is a set of sensor nodes capable of wireless communication and data processing, which are deployed over an area to gain sensing data from the environment. Sensor networks can be used for environmental monitoring, asset management, surveillance, etc. Because these networks do not necessitate wired infrastructure, they are easy to deploy and efficient at retrieving data in terms of the hardware and labor required. In addition to the general QoS attributes of communication systems such as timeliness or reliability, sensor networks are concerned with energy efficiency because they mostly operate in an unattended mode and, hence, the remaining energy determines the network lifetime. Thanks to the advances in embedded platforms and power saving technologies, wireless sensor networks are becoming popular not only in research institutes but also in general society [1,2]. Wireless sensor networks (WSN) have been developing rapidly in the last years and they promise to be one of the basic infrastructures to support ambient intelligence. Currently, sensor nodes are promoted to be used in a wide spectrum of applications such as in military, healthcare and environmental applications [4,5], where they manage highly sensitive information. This means that compromisation of

data is unacceptable in these mission critical environments since they are depended on timely and reliable information to provide their services. In WSNs, there are many risks [3] that can compromise the network and its data. Therefore, security in critical WSNs infrastructures becomes essential and must be addressed in order to protect the network and its data.

In computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (also known as evolutionary computation) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination). The process of a genetic algorithm usually begins with a randomly selected population of chromosomes. These chromosomes are representations of the problem to be solved. According to the attributes of the problem, different positions of each chromosome are encoded as bits, characters, or numbers. These positions are sometimes referred to as genes and are changed randomly within a range during evolution. The set of chromosomes during a stage of evolution are called a population. An evaluation function is used to calculate the "goodness" of each chromosome. During evaluation, two basic operators, crossover and mutation, are used to simulate the natural reproduction and mutation of species. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes. A typical genetic algorithm requires two things to be defined:

1. A genetic representation of the solution domain,
2. A fitness function to evaluate the solution domain.

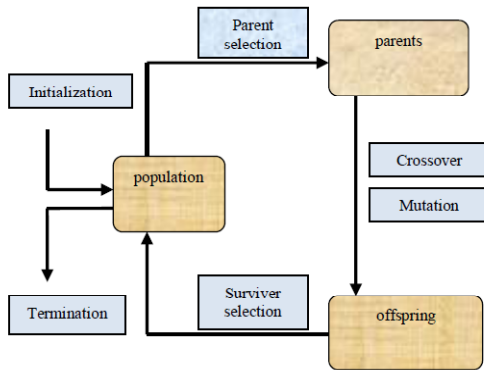


Fig. 1. The general scheme of an evolutionary algorithm

And we can draw a GA's flowchart for its iterative process like Fig1. When a GA is used for problem-solving, three factors will have impact on the effectiveness of the algorithm, they are: 1) the selection of fitness function; 2) the representation of individuals; and 3) the values of the GA parameters. GA is commonly used in applications where search space is huge and the precise results are not very important [8]. As GA is relatively computation intensive, it is executed only at the base station

so genetic algorithms can be used to healing coverage hole in wireless sensor networks. The reminder of the paper is organized as follows:

Section 2 provide a brief review of some costly related works. The proposed protocol is describe in section 3. Concluding remarks appear in section 4.

## 2 Related Work

We now review related work in geographic routing protocols and describe the essentials of geographic routing for context. There is a broad literature on geographic routing: from initial sketches suggesting routing using position information [19]; to the flooding based and backtracking based proposals for guaranteed delivery following Finn’s initial work [14], including f-GEDIR [15], SPEED [16], PAGER-M [18], and DFS [17]; to the first face routing based detailed proposals, including GFG GPSR [13], and the GOAFR+ family of algorithms [12]; to refinements of these proposals for efficiency [9], using virtual coordinates [11], and even geographic routing without location information [10]. We first describe the shared characteristics of the combined geographic routing algorithms with guaranteed delivery, and here after refer to this family of algorithms simply as geographic routing.

## 3 Proposing New Method

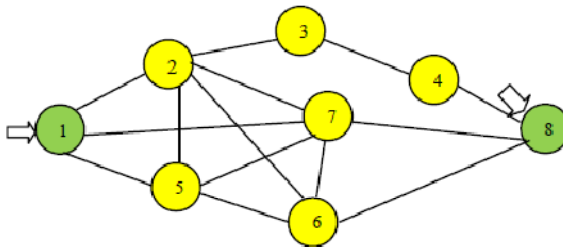
By paying attention to previous routing protocol in wsn we realize that in some of algorithms the main criteria is distance

And in some of others the criteria is energy. In our work we have two parameter energy and distance, so our proposed method have both of criteria.

### A. Nodes deployment schema

Our schema of nodes shown in Fig. 1

In this figure source and destination sensors is shown by green color and forwarding sensors is determined by yellow color.



**Fig. 2.** Geographical view of forwarding packets

In the some traditional methods just distance criteria is consider in routing protocols. But in this way because of unbalancing routing some sensors has a higher

traffic that called bottleneck that is shown in fig2. So to balance of energy consumption and lower traffic we propose the genetic algorithm method.

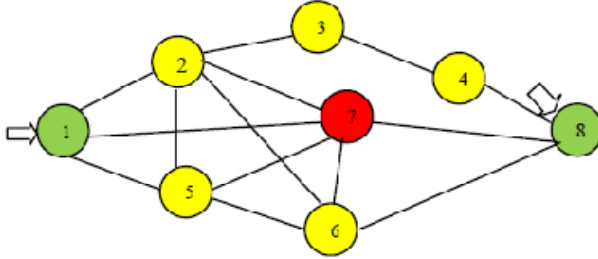


Fig. 3. Bottleneck problem determine by red sensor

B. Chromosome Representation

chromosome represents nodes and their activity in routing, the first rows determines number of nodes and the second rows identifies the state of node activities that participant in routing It is shown at Fig.1.

1	2	3	4	5	6	7	8
2	1	0	0	0	0	1	3

Fig. 4. Representation of chromosome

In our chromosome we use the following fitness function by consider the distance and remaining sensors energy that determined in below relations.

$$\text{Fitness} = \sum_{i=1}^{i=n} (\alpha E + \beta D) \tag{1}$$

$$E = \sum_{i=1}^{i=n-2} S / n - 2 \tag{2}$$

$$D = \sum_{i=1}^{i=n} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

$$\alpha + \beta = 1 \tag{4}$$

C. Parent Selection and survivor selection Mechanism

The role parent selection is to distinguish among individuals based on their quality, in particular, to allow the better individuals to become parents of the next generation .An individual is a parent if it has been selected to undergo variation in order to create offspring .Together with the survivor selection mechanism, parent selection is responsible for pushing quality improvement. It is probabilistic. High-quality individuals get a higher chance to become parents than those with low quality. Nevertheless, low-quality individuals are often given a small, but positive chance.

D. Variation Operator

The role of variation operators is to create new individuals from old ones. In the corresponding phenotype space this amounts to generating new candidate solutions. From the generate-and-test search perspective, variation operators perform the "generate" step. Variations operators in genetic algorithm are divided into two types based on their aim. There are crossover and mutation so we use them as a follow.

E. Two-Point Crossover

A single crossover point on both parents' organism strings is selected. All data beyond that point in either organism string is swapped between the two parent organisms. The resulting organisms are the children. Fig .2 represent one-point crossover.

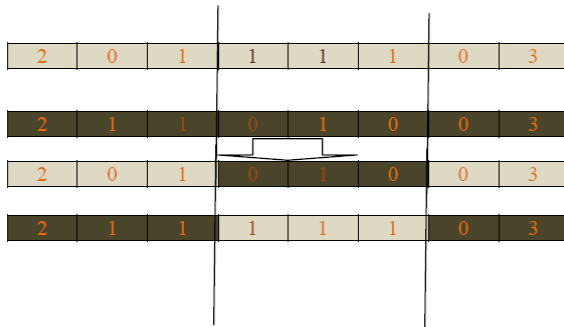


Fig. 5. Two-point crossover

F. Mutation for Binary Representations

Common mutation operator used for binary encodings considers each gene separately and allows each bit to exchange by others with a small probability  $p_m$ . In Fig.3 this is illustrated for the case where the second and fourth random values generated are less than the bitwise mutation rate  $p_m$ .

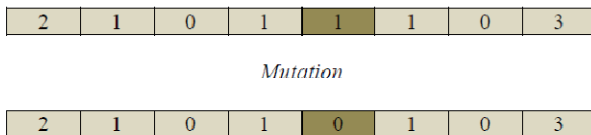


Fig. 6. Flip flop mutation

A number of studies and recommendations have been made for the choice of suitable values for the bitwise mutation rate  $p_m$ . And it is worth noting at the outset that the most suitable choice to use depends on the desired outcome. For example: does the application require a population in which all members have high fitness, or simply that one highly fit individual is found? However most binary

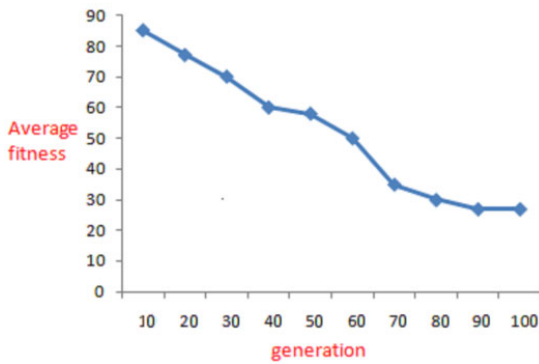


coded GAs use mutation rates in a range such that on average between one gene per generation and one gene per offspring is mutated.

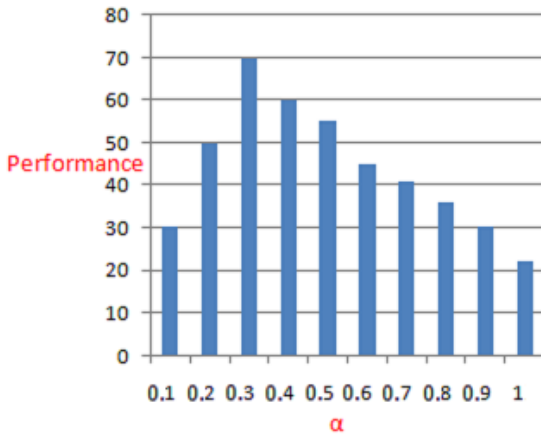
The simulation parameters are shown in Table 1.

**Table 1.** Simulation parameters

Parameter	Value
Numbers of sensor	8
$P_c$	80%
$P_m$	5%
Parent Selection Mechanism	Elitism
Survivor Selection Mechanism	Elitism
Number of Generation	40



**Fig. 7.** Generation and their average fitness



**Fig. 8.** Effect of a criteria and their performance

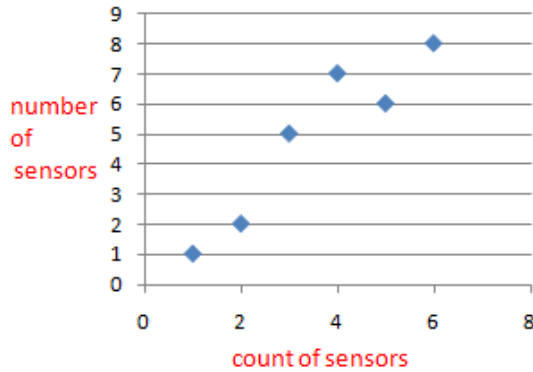


Fig. 9. The best chromosome active gens

## 4 Conclusion

In this paper, we proposed a genetic algorithm to improve routing problem that cause to prolong lifetime. In our proposed algorithm we use two criteria ,distance between sensors and resident energy. This algorithm cause to balance routing .We simulate proposed algorithm by MATLAB software. Simulation results show that when we use good coefficient of our parameters the life time of wireless sensor network is improvement.

## 5 Future Work

We will consider other parameters like priorities packets and heterogenous nodes in deployment environment.

## References

1. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Computer Networks: The International Journal of Computer and Telecommunications Networking* 38(4), 393–422 (2002)
2. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the world with wireless sensor networks. In: *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (May 2001)
3. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and countermeasures. In: *IEEE International Workshop on Sensor Network Protocols and Applications*, May 2003, pp. 113–127 (2003)
4. Romer, K., Mattern, F.: The design space of wireless sensor networks. *IEEE Wireless Communications* 11(6), 54–61 (2004)
5. Garcia-Hernandez, C.F., Ibarquengoytia-Gonzalez, P.H., Garcia- Hernandez, J., Perez-Diaz, J.A.: Wireless sensor networks and application: a survey. *International Journal of Computer Science and Network Security (IJCSNS)* 7(3) (2007)

6. Finn, G.G.: Routing and addressing problems in large metropolitan-scale internetworks, Technical Report ISI Research Report ISI/RR-87-180. University of Southern California (March 1987)
7. Stojmenovic, I., Lin, X.: Loop-free hybrid single-path/flooding routing algorithms with guaranteed delivery for wireless networks. *IEEE Transactions on Parallel and Distributed Systems* 12(10) (2001)
8. Karp, B., Kung, H.: GPSR: greedy perimeter stateless routing for wireless network. In: *Proc. of the 6th Annual ACM/IEEE Int'l. Conf. on Mobile Computing and Networking (Mobicom)*, pp. 243–254 (2000)
9. Gao, J., Guibas, L., Hershberger, J., Zhang, L., Zhu, A.: Geometric spanner for routing in mobile networks. In: *Proc. of the ACM Int'l. Symposium on Mobile Ad Hoc Networking and Computing*, pp. 45–55 (2001)
10. Rao, A., Ratnasamy, S., Papadimitriou, C., Shenker, S., Stoica, I.: Geographic routing without location information. In: *Proc. of the ACM/IEEE Int'l. Conf. on Mobile Computing and Networking, Mobicom (2003)*
11. Newsome, J., Song, D.: GEM: graph embedding for routing and data-centric storage in sensor networks without geographic information. In: *Proc. of the 1st Int'l. Conf. on Embedded Networked Sensor Systems, Sensys (2003)*
12. Kuhn, F., Wattenhofer, R., Zollinger, A.: Worst-case optimal and average- case efficient geometric mobile ad-hoc routing. In: *Proc. of the 4th Int'l. Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc)*, pp. 267–278 (2003)
13. Gao, J., Guibas, L., Hershberger, J., Zhang, L., Zhu, A.: Geometric spanner for routing in mobile networks. In: *Proc. of the ACM Int'l. Symposium on Mobile Ad Hoc Networking and Computing*, pp. 45–55 (2001)
14. Karp, B., Kung, H.: GPSR: greedy perimeter stateless routing for wireless networks. In: *Proc. of the 6th Annual ACM/IEEE Int'l. Conf. on Mobile Computing and Networking (Mobicom)*, pp. 243–254 (2000)
15. Datta, S., Stojmenovic, I., Wu, J.: Internal node and shortcut based routing with guaranteed delivery in wireless networks. In: *Proc. of the IEEE Int'l. Conf. On Distributed Computing and Systems (Wireless Networks and Mobile Computing Workshop)*, April 2001, 461 p (2001)
16. He, T., Stankovic, J.A., Lu, C., Abdelzaher, T.: SPEED: a stateless protocol for real-time communications in sensor networks. In: *Proc. of the 23rd Int'l Conf. on Distributed Computing Systems (2003)*
17. Stojmenovic, I., Russell, M., Vukojevic, B.: Depth first search and location based localized routing and QoS routing in wireless network. In: *Proc. of the Int'l. Conf. on Parallel Processing, ICPP 2000 (2000)*
18. Zou, L., Xiong, Z., Lu, M.: PAGER-M: a novel location-based routing protocol for mobile sensor networks. In: *Proc. of the 1st Int'l. Workshop on Broadband Wireless Services and Applications, San Jose, CA (2004)*
19. Kleinrock, L., Takagi, H.: Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transactions on Communications* 32(3), 246–257 (1984)
20. Krishnamachari, B., Estrin, D., Wicker, S.: The impact of data aggregation in wireless sensor networks. In: *Proc. of DEBS 2002 (2002)*

# Modern Education Technology and the Transformation of Teacher Role

Canlong Wu<sup>1,3</sup> and Shuying Zhan<sup>2,3</sup>

<sup>1</sup> Education Technology Center

<sup>2</sup> School of Communications and Electrics

<sup>3</sup> Jiangxi Scientific and Technical Normal University, Nanchang, Jiangxi Province, China  
wucl@jxstnu.edu.cn, wwccll2@sohu.com

**Abstract.** The widely using of modern education technology is a very important sign about education modernization. It means not everything but lines with the modern understanding of education law, updates teaching models, contents, methods and etc, and the most important thing is that modern education technology will give a deep and wide impaction on traditional teacher role. The teacher can not be adapt to the need of the development of education in this information era if he can't reaching the changing of his own role soon. So, teacher should seize the opportunity, discard the passed concept of education, hold the inner rules of education development and the characteristics of modern ear, set up righteous concept of education and talent, improve the information literacy and the capabilities of using modern education technology, strengthen the researching in education science. Only in this way can we meet the requirements of modern educational technology.

**Keywords:** Education technology, Transformation of teacher role, School education.

## 1 Introduction

The education in the 21st century will be a new type of modern education which differs from the old one in several aspects such as idea, thoughts, methods and actions, especially with the pushing of multimedia and network, education technology has acquired fast development. The developing and applying of education technology will cause a deep and extensive effect on education thoughts, idea, teaching contents, methods, styles of organization and teacher role, and it raises a new demand on the changing of teacher role. With the deep researching and exploring of these things will provide important reference for the modernization of school education.

## 2 Modern Education Technology Has Deep Meaning on Realizing Education Modernization

The so-called modern education technology is just by using modern education theory and modern information technology through the procedure of teaching and learning

and the designing, developing, applying, managing, evaluating of teaching resources, in order to realize optimized teaching theory and practice. The extensive applying of modern education technology is an important symbol of education modernization. Compared with the traditional education technology, the active function is more apparent, it not only conforms to modern education of cognitive laws, renews teaching mode, teaching content, teaching method and means etc, but has full impact and challenges to the traditional teaching theory and practice and will eventually lead to education integrity and fundamental reform.

### **Modern Education Technology Advancing the Traditional Teaching Mode Transformation Classroom Teaching the Basic Pattern of the School Teaching**

Since the 17th century, various types and kinds of schools all over the world have adopted the same forms to organize teaching. It may be a scientific, efficient, popular education and the advantages of cultivating talents, but along with the progress of the society, the development of science, its drawbacks were exposed more and more, such as not in favor of the student's main role in accordance with their aptitude, and was unfavorable to the detriment of cultivating students' innovative spirit and practice ability, etc. And the using of multimedia, network, and modern information technology in school teaching can remedy defects in the traditional teaching mode. Modern education technology teaching system, is a teaching environment and system that one can undertake individualized autonomous learning, and at the same time, it can form the coordination of teaching environment and systems, this interactive function, can provide students with different kinds of choices. In interactive learning environment, students can select the study content and study ways suitable for personal characteristics according to their study interest and level, which can maximum exert students' enthusiasm and initiative. And the multimedia teaching systems that combined with the Internet technology, not only embodies the students' cognitive main body role, but also makes communication between students and students, students and teachers has no limitations to each other, realize the cooperative learning of free discussing. This not only helps knowledge expansion and difficult problems solved, and still can cultivate students' cooperation spirit and communication skills.

### **The Applying of Modern Education Technology Sparks the Updating of Teaching Content and Teaching Methods**

The teaching contents is one of the basic elements constitute the teaching process, is the important condition to realize education purpose and the goal of cultivating, it updates with the progress of the society and the development of science and technology. As the core of the modern information technology, today's multimedia and network technology come into all teaching fields with amazing speed and teaching content, method and means etc. is updated. With the development of education technology, from the external to the internal structure teaching content has a corresponding change. Traditional school textbooks should basically be descriptive text, science & engineering courses are complemented with graph, form, physical model, can not perform teaching content by using audio, images, text, animation in

all. The traditional written materials and auxiliary materials are linear knowledge structure, and strictly apply the rule of from easy to difficult, from near to far, from the concrete to the abstract principle in the organization and arrangement of teaching content. Students can only gain knowledge step by step under the teaching of teacher. Students are dependent to teachers largely. Modern education technology is a kind of method closing to human cognitive characteristics to organize, display teaching content and construct knowledge structure. In the study, students can break through the restrictions of time and space, micro and macro, history and reality. By using simulation, imitation, visualizations, reality, development of perceived things and the formation, process of changes can be illustrated vividly with the methods of students' cognition of the object, It makes theorem lifelike. These diverse flexible vivid teaching methods and means have greatly reduced the difficulty of the students' knowledge and can make the teaching process and the teaching effect to reach optimized state.

### **Modern Education Technology Is Able to Cultivate Students' Innovative Spirit and Practice Ability**

The fundamental purpose of education informatization is to promote the quality education, realize the cultivation of innovation talents. The development of future society mainly depends on the national wisdom, rather than the resources. Only when a person dares to develop innovative spirit, can discover something. With invention and creativity, society can develop. Advantages of modern education technology mainly are shown in the diversity of resource, the interactivity in learning process, the independency in running and etc. If our teachers can utilize these advantages in teaching process, we can arouse the enthusiasm of students' study and, arouse the students' innovation consciousness and cultivate students' divergent thinking and innovative thinking ability, form good habits of thinking.

The innovation consciousness of students should be trained in practice, and enhanced on the basis of raising practice ability. While in conditions that education technology is widely used in teaching, the teacher must train and cultivate the students with the skills of information acquisition, processing, analysis, expression and communication. So that students can continuously improve the operational skills, and use information technology tools, create special works in information technology applying and practicing.

## **3 Teacher Role Conversion in Modern Education Technology Conditions**

For a long time, people compare the teacher to "engineers of human souls", one who preach knowledge and reassure the human culture, scientific knowledge and social ideas, morality. Teacher's occupation is holy in people's mind and teacher's position is important. But with the development and application of modern education technology, the role of teacher will change. Teachers will no longer be the only source of student that can obtain knowledge; the teacher's authority will gradually fade. However, the changing of the role does not mean that their social status become low and

insignificant. In fact, the requirements for teachers become stricter. Although the application of information technology has brought new challenges for teachers in teaching process, make significant changes in teaching process about the teacher role, what should be emphasized is that the information technology can never replace the role of teachers. Teachers are, and will always be the indispensable important factors that consist of teaching process.

The key of realizing education modernization lies in teachers. The new teacher's function is diversified. First, teachers' role has changed from the traditional teaching center into student-center. When students' learning objective is ambiguous, teacher should guide students set suitable learning goals, When students' learning methods are not proper and when it will affect learning effects, teachers should teach students learning strategy and scientific methods which can help them obtain effective information. When students are not interested in learning, the teacher should induce and stimulate the students' learning motivation, cultivate students' interest, arouse students' enthusiasm. When students meet difficulties in learning, the teacher should counsel them promptly, help students solve the obstacles. When students are polluted by bad information, teachers should teach students to choose and absorb correct useful information, improve the students' political keen and discriminate the bad or right information, resist the temptation brought by bad information and new technology. Second, modern education technology has replaced part of a teacher's function, the teacher will be liberated from heavy teaching work, so they can have more time and energy and engage in the course designing, developing, software making and researching in scientific education, etc. The role of teacher will change from "JiaoShuJiang" to "researcher" and become a respected educator.

#### **4 Strategy of Realizing the Conversion of Teacher Role**

The widely used modern education technology in school, are not only benefit to deepen teaching reform and improve teaching quality and efficiency, but also play a positive role in promoting the quality education and training to adapt to the requirement of creativity in information age. But it put full impact and severe challenges on the traditional teacher role. We should seize the opportunity of development, research actively and explore strategies, take effective measures to promote the conversion of teachers role.

##### **Change Education Concept**

Change education concept is the prerequisite to realize education modernization. All kinds of school should update education concept under the guidance of scientific development, sets up the correct education concept.

First of all, we must transform the traditional concept of knowledge and talent. Secondly, transform traditional view about teachers and students. By applying modern education technology into teaching, the authority position of a teacher as in original teaching has changed to a position which will guide students to study and be a partner of them. In conditions that the new technique is applied to teaching, teachers and students' communication channels and space will be broadened more, the

communication mode and means will be various and flexible, which will help to build democratic, equal and harmonious society by coordinating the relationship between teachers and students; Again, we must transform the traditional teaching management and teaching assessment view. Teaching evaluation means school-related department or relevant personnel make the value judgment according to a certain value standard or value system, and adopt corresponding methods and means of teaching. Scientific evaluation will make good guidance to education and teaching activities. The traditional teaching evaluation puts emphasis on knowledge, transverse comparison, less cares ability, longitudinal comparison. Whereas the current teaching evaluation is diversified, must examine knowledge, abilities and skills. We need a quantitative evaluation, also should have a qualitative evaluation. We need a self-assessment, also should have others' review.

### **To Improve Teachers' Information Literacy and Ability of Using Modern Education Technology**

The development of science and technology brings the advanced technology and equipment to school education. Facing the new education technology, how to make teachers adapt to new media, control the new media, is directly related to the education modernization and new qualified talents training. Thus, it has become an urgent and difficult task in school reforming and developing about improving teachers' information literacy and ability of using modern education technology.

At present, many schools have bought some modern education equipment, some schools have campus network, but the utilization rate is not high. There are two reasons. One is the recognition and the attitude. Some teachers content to traditional teaching methods such as chalk with blackboard, textbooks with notes. They think the advanced teaching media is detrimental to individuality, some teachers, especially the olders, are afraid of facing new teaching media, some of them are not familiar with new technology, they are afraid of losing face not class in front of the students if they will get stuck. Second, the school pay less attention to the training of improving teachers' information literacy. As it is known to all, the enhancement of teachers' information literacy and information technology using ability not only helps to convert the role of teachers, make teachers undertake the historical task of modern education better, but also helps to improve the utilization of new technology equipment in school, push the process of school teaching and management information, improve teaching efficiency.

Therefore, it is the urgent task of strengthening the education modernization about teachers in current schools. We must pay close attention to it and do it better.

### **Advocate Teachers' Scientific Research in Education Vigorously**

Scientific research in education is to adapt to the need of education practice, it a kind of activity about studying education practice problems with scientific attitude and approaches, refining education experience and probing education law.

Many people think only a university professor or staff in research organization has the ability to do research, and teachers in foundation education have no conditions of engaging in scientific research, therefore neglect education research activities.



At present, facing the new technical challenges, it is difficult to adapt to the demand of information society only by past experience of teaching, it is also hard to achieve education reforms truly. Education research means deepening education reform, is the intermediary and bridges that transform the teacher's education literacy into education effect. An outstanding teacher should seize the opportunity of applying education technology in school, combine teaching and researching together organically, select some meaningful projects in the theory and practice, research and explore various education phenomenon and the rule of education in the new technology conditions, make scientific education walks in front of practical workers, especially, some forward-looking research results will have more value of guiding education work.

Teaching and research are mutual supplement and reciprocally stimulative and mutual improved. Practice has proved that those who ready to conduct research, their knowledge can constantly be opened, capacity of developing teaching software be improved continuously and thinking ability, innovation ability and self-educated ability, scientific thesis writing ability which reflect reserching ability also be improved to a new height. It is so attractive, why should not do?

## References

1. Liu, L.-Z.: The reform and enlightenment of international talents training mode in developed countries. *Modern Education Science* 7, 12 (2009)
2. Zhou, R.: Informatization of teaching equipment based on multivariate theories. *Journal of Hefei University of Technology*, 12–17 (March 2007)
3. Li, M.: Some education problems in the informatization education of middle and primary schools. *Teaching Instrument and Experimentation*, 55 (July 2003)

# Network Security Risk Assessment Model and Method Based on Situation Awareness and CORAS

Yong Qi, Yan Wang, and Qianmu Li

School of Computer, Nanjing University of Science & Technology, Nanjing, 210094, China  
{qyong,qianmu}@mail.njust.edu.cn, wyanxz@126.com

**Abstract.** Based on Network Security Situation Awareness framework, combined with DS Theory and CORAS, this paper proposes a Network Security Situation Awareness Risk Assessment Model. Then, we take Low Earth Orbit (LEO) Satellite Communications Network as an example to make a simulation and assessment. Through the experiment, the proposed model and method are verified to be valid.

**Keywords:** DS Theory, CORAS, Situation Awareness, Network Security, Risk Assessment.

This paper based on Network Security Situation Awareness framework, combined with DS Theory and CORAS, proposes a Network Security Situation Awareness Risk Assessment Model. Then, we take Low Earth Orbit (LEO) Satellite Communications Network as an example to make a simulation and assessment. Through the experiment, the proposed model and method are verified to be valid.

## 1 Overview of the Related Theory

### 1.1 Situation Awareness of Network Security

The concept of situation awareness (SA)[1] comes from the research on human factors of the aerospace field. In 1988, Endsley defined SA as "Perception of the element attributes for a certain time, within a certain space, namely, the understanding of the meaning and the forecast of the recent state". In 1999, Tim Bass inspired by the concept of Situation Awareness which named Air Traffic Control (ATC), proposed the Situation Awareness framework in the field of network security. Network Security Situation means that the global security state of the monitored network, the attack of a time window and the impact of the overall goal of network security. NSSA described the real-time risk information of the entire network, and global security risk Situation, it is the research focus of the field of current network security assessment. Many scholars study the model of NSSA, and focused on the study of framework model[2], such as the framework of multi-sensor data fusion proposed by Tim Bass, the NSSA framework model based on NetFlow proposed by Xiaoxin Yin, the distributed NSSA framework model proposed by

Stephen G Batsell and so on. While these framework is mainly about the qualitative analysis, no precise quantitative methods of risk. Considered the related study in domestic and foreign, the conceptual model of NSSA framework as shown in Fig. 1.1 [3], which divided into three levels: ① Situation Recognition, ② Situation Assessment, ③ Situation Forecasts. The first level is situation recognition, which is the basic of Situation Awareness, and it uses mature technology to identify network security situation information from the massive multi-source heterogeneous data, then, transformed into a comprehensible format (such as XML), prepared for situation assessment. The second level is situation assessment, it is the core of situation awareness, and it is a dynamic, real-time understanding of the process of network security situation, Through awareness safety events, and determine the relationship between them, generate security situation graph. The third level is situation forecasts, it is the process which according to the previous and current network security situations, get the network security posture, then according to real-time network information to determine trends of security.

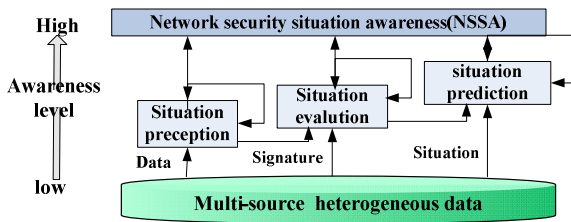


Fig. 1.1. The conceptual model of network security situation awareness

### 1.2 Risk Analysis of CORAS

To compensate for the shortcomings of NSSA framework that does not quantify, we introduced the risk analysis framework based on CORAS method to analysis all stages of situational awareness. CORAS framework[4] is a platform for risk analysis of security critical systems which completed by Germany, Greece, Britain, Norway in 2003.

The platform combined the risk analysis techniques and the UML-based system modeling techniques together, including terminology, libraries, methodologies and tools four parts. Terminology defined the concept related to the framework, gave unified naming standards; methodology gave techniques about risk analysis, the process followed and the methods of describes, provide a theoretical basis for analysis; knowledge libraries provide a priori knowledge for risk analysis, to assess the results of risk analysis data which contains the storage per times to improve analysis accuracy and efficiency; toolset provides calculation tools of methods, and used to implement knowledge libraries.

### 1.3 DS Evidence Theory

DS evidence theory[5] was proposed by Arthur Dempster in 1967, and then further developed by Glenn Shafer, formed the evidence theory based on belief function.

**(1) Mass Function**

Suppose that a decision problem was composed by n decision direction or the state, formed a mutually exclusive set of all equipment, denoted by  $\{a_1, a_2, \dots, a_n\}$ , we call the entire collection as recognition framework, denoted by  $\Theta$ . The basic probability assignment function BPA (mass function), which is the interval mapping to identify the power set  $2^\Theta$  of the framework  $\Theta$  to  $[0,1]$ ,  $m(A)$  is called basic probability assignment BPA, or m value, said the credibility of the evidence of A, Besides, it can also be described as the evidence of support to the occurrence of A. An evidence assigned the m values to all subsets of the framework was added up to 1:

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{1}$$

Where A stands for all the subsets of the framework  $\Theta$ , and  $m(\phi) = 0$ , that the m value of empty set is 0.

**(2) Dempster Combination Rule**

Dempster rule is the basis rule of DS theory, which is used for the integration of two or more evidence. We record the mass function as  $m_1, m_2, \dots, m_s$ , which identify s independent evidences of the framework  $\Theta$ , for  $\forall A \subseteq \Theta$ , if the m value which is the total synthesis of  $m_1, m_2, \dots, m_s$  recorded as  $m(A)$ , the combination rules of  $m_1, m_2, \dots, m_s$  is:

$$m(A) = m_1(A) \oplus \dots \oplus m_s(A) = \begin{cases} 0 & A = \phi \\ \frac{1}{K} \sum_{\bigcap_{i=1}^s A_i = A} [m_1(A_1) \bullet \dots \bullet m_s(A_s)] & A \neq \phi \end{cases} \tag{2}$$

where  $K = \sum_{\bigcap_{i=1}^s A_i \neq \phi} [m_1(A_1) \bullet \dots \bullet m_s(A_s)] = 1 - \sum_{\bigcap_{i=1}^s A_i = \phi} [m_1(A_1) \bullet \dots \bullet m_s(A_s)]$

The main features of the evidence theory: (1) Mass function includes "uncertain" information, and in Dempster combination, these information is kept, and integrated as a combined result. (2) DS theory is not only able to assign trust for individual element which in the identify spatial, but also able to assign trust for its subset. Using DS theory to integrate multiple sensors can effectively improve the recognition accuracy of the threat situation.

**2 Network Security Situation Awareness Risk Assessment Model**

In this paper, we take LEO satellite communication network as the example, combined with the DS evidence theory, CORAS risk analysis and probabilistic analysis, security association trees and other methods, proposed the Network Security Situation Awareness Risk Assessment Model, as shown in Fig.2.1:

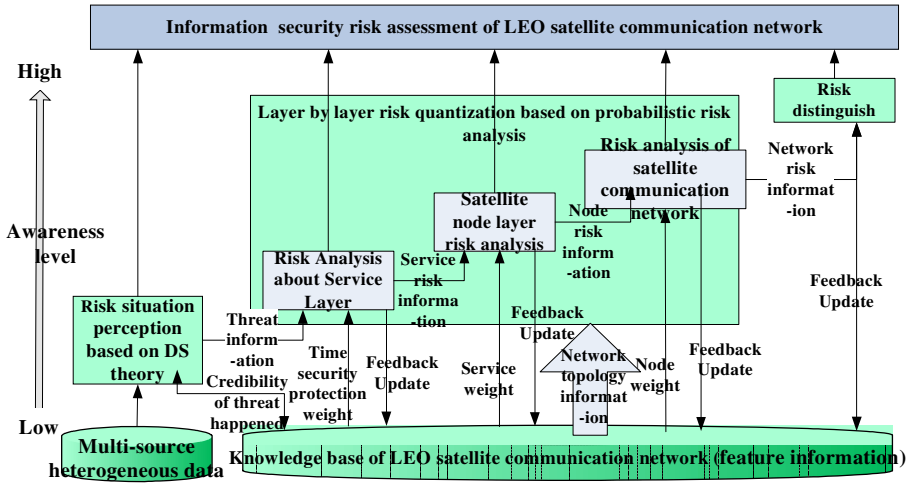


Fig. 2.1. CORAS Risk Assessment Model that based on Network Security Situation Awareness

From bottom to top, LEO satellite communication network is divided into three risk levels: the risk of security services, the risk of node, and the risk of communication network. Model divides the whole assessment into three awareness stages:

(1) Threat situation recognition based on DS theory. We take firewall, intrusion prevention system and other network security tools as sensor, to get the potential attacks information of network (including the type of attack, the possibility and other information) that may exist, and then combine attack information perceived by various of sensors based on DS evidence theory, get the occurrence threats and the credibility of whether the threat will happen, take it as the input of the second stage, therefore, we don't need to take care of thousands of warning data, improve the efficiency of network management and control.

(2) Layer by layer risk quantification based on probabilistic risk analysis. For the threats identified in phase 1, using methods that based on probability analysis and security association tree, do analysis to quantify risk of service layer, risk of satellite node layer and risk of network layer. And in the analysis process, we take the weight of time, the importance of security services on satellite nodes, the importance of satellite nodes on in the network, and other information into account, in order to increase objectivity, and besides, we also take the protect role which security mechanism plays, and together with the situation that risk spread along the nodes into account.

(3) Risk discrimination. According to the result of risk analysis in the second phase, we do discrimination the risk of the whole network, including the discrimination of the most risky network nodes, and the type of risk network most likely to face.

In each stage of awareness, we take the risk analysis method based on CORAS, give a specific risk analysis method, knowledge libraries, terminology, implementation tools. Besides, we introduce the feedback and update mechanism, adjust and update repository based on the actual situation in the assessment process.

### 3 Network Security Risk Assessment

#### 3.1 Network Threat Situation Recognition Based on DS Theory

Framework of threat situation recognition:

- (1) Terms: Threat, threat situations, sensor, identification framework, mass function, trust;
- (2) Methodology: DS evidence theory;
- (3) Tools: Use firewall, intrusion prevention system, antivirus software and security detection technology as a sensor;
- (4) Knowledge base: Network flutter model diagram, sensor perceived credibility of various threats occurred (Tab. 3.1).

Definition 1: Threat  $Threat = (T\_name, T\_toSA, T\_time, T\_probability)$ , which said that  $T\_name$  is the name of threat,  $T\_toSA$  is the security services that affected,  $T\_time$  is time that threat occurred,  $T\_probability$  is the possibility that the threat occurred.

Definition 2: in accordance with the process type of network application layer, threat can be divided into: SNMP Threat, RPC Threat, FTP Threat, HTTP Threat, TELNET Threat and DNS Threat.

Definition 3: Use firewall (FW), intrusion prevention system (IPS), antivirus software (AV), and security detection technology (SIT) as the possible threats from a multi-source sensor networks data,

Identification framework  $\Theta = \{threat\ exist\ threat\ does\ not\ exist\}$ , sensor  $S = \{S_1, S_2, \dots, S_s\}$ , we take sensor  $S_i$  on the threat  $Threat_j$  that indeed reported to explain  $M^{ij}$  which is said to be the credibility, that  $m(threatened) = M_{ij}$  = indeed reported rates, we take the false alarm rate of the threat  $Threat_j$  of sensor  $S_i$  to describe the credibility  $\tilde{M}_{ij}$  that sensor  $S_i$  perceived  $Threat_j$  is not exist,  $m(threat\ does\ not\ exist) = \tilde{M}_{ij}$  = False alarm rate, credibility is coming from the experiment, stored in the knowledge base. The credibility that each sensor perceives a threat is on the presence or absence as shown in Tab. 3.1, in which the items not filled means that the threat of the corresponding column item can not be perceived:

**Table 3.1.** The credibility that FW, IPS and other sensors on the occurrence of SNMP, RPC and other threats

Cred- -ibility	SNMP Threats		RPC Threats		FTP Threats		HTTP Threats		TELNETThreats		DNSThreats	
	Exist	Does not exist	Exist	Does not exist	Exist	Does not exist	Exist	Does not exist	Exist	Does not exist	Exist	Does not exist
FW	0.4	0.6	0.4	0.6	0.5	0.5	0.5	0.5	0.6	0.4	0.7	0.3
IPS	0.5	0.5	0.7	0.3	0.6	0.4	0.8	0.2	0.7	0.3	0.8	0.2
A-V	0.7	0.3	0.6	0.4	0.5	0.5	0.5	0.5				
SIT	0.7	0.3	0.6	0.4	0.6	0.4	0.6	0.4	0.8	0.2	0.5	0.5

The steps about the integration of threat information based on DS theory is shown as follows:

Record the information of a threat that perceived by a sensor. To Threat<sub>j</sub>, we get the evidence from whether the threat is exist or not perceived by sensor S<sub>i</sub>, denoted by m<sub>i</sub>, and s sensors can form s evidence, and the identification framework Θ = {Threat<sub>j</sub> exist, Threat<sub>j</sub> does not exist}. If S<sub>i</sub> perceived Threat<sub>j</sub>, m<sub>i</sub>(Threat<sub>j</sub> exist)=M<sub>ij</sub>, abbreviated as m<sub>i</sub>(Threat<sub>j</sub>)=M<sub>ij</sub>; m<sub>i</sub>(Threat<sub>j</sub> does not exist) =  $\overline{M_{ij}}$  abbreviated as  $m_i(\overline{Threat_j}) = \overline{M_{ij}}$ . If S<sub>i</sub> does not perceived Threat<sub>j</sub>, the value of m<sub>i</sub>(Threat<sub>j</sub> exist) will not be assigned. A basic probability assignment value will be assigned to Threat<sub>j</sub> by each evidence, or not assigned, to get accurate results.

(2) For Threat<sub>j</sub>, we merge the information about threat in s evidence of using evidence theory method, Threat<sub>j</sub> that perceived refers to a single element Threat<sub>j</sub>, (1.2) can be reduced to:

$$\begin{aligned}
 m(Threat_j) &= m_1(Threat_j) \oplus \dots \oplus m_s(Threat_j) \\
 &= \frac{1}{K} \cdot [m_1(Threat_j) \bullet \dots \bullet m_s(Threat_j)]
 \end{aligned}
 \tag{3}$$

Besides,

$$K = \sum_{\bigcap_{i=1}^s A_i \neq \emptyset} [m_1(A_1) \bullet \dots \bullet m_s(A_s)] = m_1(Threat_j) \bullet \dots \bullet m_s(Threat_j) + m_1(\overline{Threat_j}) \bullet \dots \bullet m_s(\overline{Threat_j})$$

then  $Threat_j T\_propability = m(Threat_j)$

(3) According to step (2), we can get the credibilitied of all threats occurred, and take them as the input of the second phase.

### 3.2 Layer by Layer Risk Quantization Based on Probabilistic Risk Analysis

The framework of Layer by layer risk quantization:

(1)Terms: Service risk, direct threats, indirect threats, the access associated graph, the access associated tree, node risk, network risk;

(2)Methodology: Probabilistic risk analysis, the access associated tree;

(3)Knowledge base: The weight of time (T<sub>1</sub>=(0:00~8:00), T<sub>2</sub>=(8:00~18:00), T<sub>3</sub>=(18:00~24:00), in which W<sub>T1</sub>=0.11, W<sub>T2</sub>=0.67, W<sub>T3</sub>=0.22), security services set SA, the relative importance weight w<sub>sj</sub> that security services SA<sub>j</sub> on node N<sub>i</sub>, The losses SA<sub>ij</sub>-Impact of node N<sub>i</sub> that affected by the security services SA<sub>j</sub>(quantitative

criteria: 3: high, 2: middle, 1: low), the threat of harm to the security services (Tab. 3.2), The Security index  $g_{ij}$  that the security Mechanism  $M_i$  for security services  $SA_j$  (Tab. 3.3, quantitative criteria: 3: high, 2: middle, 1: low), the relative importance weight  $wn_i$  of  $N_i$ , and the access associated graph, the access associated tree between nodes.

**Table 3.2.** The threat that do harm to security services

Threat Security Service	SNMP Threat	RPC Threat	FTP Threat	HTTP Threat	TELNET hreat	DNS Threat
Availability	V	V	V	V	V	V
Integrity	V	V			V	
Confidentiality	V	V	V	V	V	
Certification	V	V	V		V	
Non-repudiation	V				V	

Note: V indicates a threat in the column may cause harm to the security services in the line.

**Table 3.3.** The protective role  $g_{ij}$  that security mechanisms played for security services

Security Services $SA_j$ Protective mechanism $M_i$	Availability $SA_1$	Integrity $SA_2$	Confidentiality $SA_3$	Certification $SA_4$	Non-repudiation $SA_5$
Encryption and key management	2	2	3	1	
Access Control	1	1	1	3	
Authentication			1	3	2
Error Control	3	3			1
Wireless Spread Spectrum	3	2	2		

The risk of the entire network[6] is decided by the risk of nodes that formed the network, the risk of node is related to the risk about its security services threats. Based on the analysis of threats about security services of each node, and based on the Probabilistic Risk Analysis(PRA) method, we do analysis on security risks of the corresponding layer from bottom to top quantitatively layer by layer(including the value of risk and the probability of the risk ), finally,do risk assessment on the entire network to obtain the total risk.

**3.2.1 Risk Analysis about Security Service Layer**

Threat that service suffered on node  $N_i$  is related to three aspects: ① Direct threat, the harm that an attacker use the vulnerability of  $N_i$  to attack it; ② Indirect threat, that node  $N_k$  have a legitimate relationship with  $N_i$  about access, and after the attacker attack  $N_k$ , then, using the relationship of  $N_k$  and  $N_i$  to access and control node  $N_i$ , and will cause harm; ③ The protective function that security mechanisms played of node itself.

Definition 4: Probability of service risk  $PS_{ij}(t)$ : the probability that threat happened to security services  $SA_j$  provided on  $N_i$  at the moment t.



Definition 5: The probability of a direct threat  $PS_{ij}^d(t)$ : the probability that a direct threat happened to security services  $SA_j$  provided on  $N_i$  at the moment  $t$ .

$$PS_{ij}^d(t) = 1 - \prod_{Thr \in Th_{ij}^t} (1 - Thr \cdot T\_probability) \tag{4}$$

Definition 6: The probability of an indirect threat  $PS_{ij}^i(t)$ : the probability that a indirect threat happened to security services  $SA_j$  provided on  $N_i$  at the moment  $t$ . When an attacker successfully exploited the corresponding weaknesses to do threat to node  $N_A$ , then, use association between node  $N_A$  and  $N_B$ , do further attack to  $N_B$ . We recorded the access relationship from  $N_A$  to  $N_B$  as  $NTC_{AB}$ ,  $NTC_{AB} = (A, B, r, p, Threat)$ ,  $r$  means that the relationship used when  $NTC_{AB}$  occurred,  $p$  is the occurrence probability of  $NTC_{AB}$ ,  $Threat$  means that threat to Node B after  $NTC_{AB}$  occurred.

Definition 7: The access associated graph  $GA = (N, E)$  is shown in Fig.3.1,  $N$  is the set of vertices, representing network nodes;  $E$  is the set of the directed edges,  $e(N_3, N_2)$  means that the access relationship  $NTC_{AB}$  between the adjacent node  $N_3$  and  $N_2$ , which is associated from the association node to another node.  $e(N_3, N_2)$ .  $r$  means that the relationship used when  $NTC_{AB}$  occurred,  $e(N_3, N_2)$ .  $p$  means that the probability of  $NTC_{AB}$ 's occurrence.

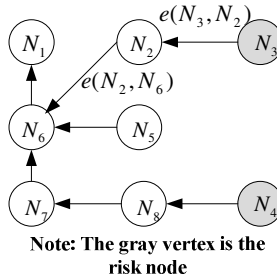


Fig. 3.1. Instance figure of the access associated graph

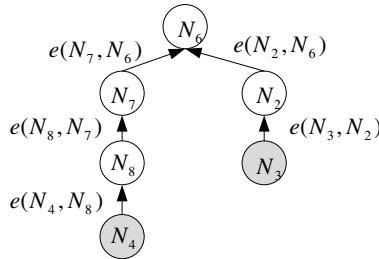


Fig. 3.2. Instance figure of the access associated tree

Definition 8: The access associated tree of node  $N_i$ : the root node is the accessed node (subject to assessment)  $N_i$  of the network; intermediate node is the node that accessed the root node through one or more layers, leaf node is the source of risk (network node that directly affected by the threat). The access associated tree of node

$N_6$  in Fig 3.2 is shown in Fig 4.1,  $N_3, N_4$  are the source of risk , the root node  $N_6$  is the node being evaluated. Up layer by layer from the leaf node level, the probability  $P_k(t)$  means that threats occurred to the comprehensive source of risk,  $e(N_k, N_l).p$  means the probability of each directed edges occurred, then, we can calculated the threat probability which come from the source of risk about the assessing node .

$$P_l(t) = P[\cup_{k=1}^n N_k \bullet e(N_k, N_l)] = 1 - \prod_{k=1}^n [1 - P_k(t) \bullet e(N_k, N_l).p] \tag{5}$$

Until the root layer is the end of operations.

$$PS_{ij}^i(t) = 1 - \prod_{k=1}^n [1 - P_k(t) \bullet e(N_k, N_l).p] \tag{6}$$

After analysis, we can get the access associated graph and the access associated tree between nodes, and stored in the repository, constantly updated, then, we can get  $PS_{ij}(t)$ :

$$PS_{ij}(t) = 1 - [1 - PS_{dij}(t)][1 - PS_{ij}(t)] \tag{7}$$

Definition 9: Risk index  $RS_{ij}(t)$  of Security service layer: the impact that threat happened to security services  $SA_j$  provided on  $N_i$  at the moment  $t$ .

$$RS_{ij}(t) = \frac{W_T(t)[PS_{ij}(t) \bullet SA_{ij} - Im\ pact]}{Guard_{ij}} \tag{8}$$

where  $W_T(t \in Ti) = w_{Ti}$ ;  $Guard_{ij} = \max_{k=1}^{mn} g_{kj}$  means the protective function that security mechanisms played for security services  $SA_j$  on  $N_i$ .

### 3.2.2 Node Layer Risk Analysis

Definition 10: Node risk index  $RN_i(t)$ : the impact that threat happened to node  $N_i$  at time  $t$ .  $RN_i(t)$  is related to the security situation of all security services provided by  $N_i$  and the relative importance on the weight of each security services , the higher the value, the greater the risk.

$$RN_i(t) = \sum_{j=1}^{ni} [ws_{ij} \bullet RS_{ij}(t)] \tag{9}$$

Where  $ws_{ij}$  means the relative importance on the weight of security service  $SA_j$  in all security services provided by  $N_i$ .

Definition 11: Node risk probability  $PN_i(t)$ : the probability that threats happened on node  $N_i$  at time  $t$ .  $PN_i(t)$  is the joint probability of all threats that security services on node  $N_i$  suffered at time  $t$ .

$$PN_i(t) = 1 - \prod_{j=1}^{an} [1 - PS_{ij}(t)] \tag{10}$$

### 3.2.3 Network Security Risk Analysis

According to the idea that from local to whole, after get risk index  $RN_i(t)$  of each network node at time  $t$ , then, the risk index of the entire network can be determined.

Definition 12: Network risk value  $R(t)$ : the total risk value of the whole LEO satellite communication network at time  $t$ .  $R(t)$  with all nodes in the network and information security status of each node in the network weights the relative importance.  $R(t)$  is related to the security situation of all security services and the relative importance on the weight of each node in the network.

$$R(t) = \sum_{i=1}^n [wn_i \bullet RN_i(t)] \tag{11}$$

Definition 13: The probability of network risk  $P(t)$ : The joint probability of threats that all security services suffered at time  $t$  in the whole network.

$$P(t) = 1 - \prod_{i=1}^n [1 - PN_i(t)] \tag{12}$$

### 3.3 Network Security Risk Identification

Through the above methods, we can get the risk value  $R(t)$  of the entire network, the probability  $P(t)$  of risk occurrence, the risk value  $RN_i(t)$  of each node and risk probability  $PN_i(t)$ , and the risk value  $RS_{ij}(t)$  about each security services of the node and risk probability  $PS_{ij}(t)$ . For security services  $SA_i$ , a max, on the one hand, according to the analysis process of the direct threat, we can determine the vulnerabilities of the node through the type of threat suffered, and then make appropriate treatment measures to deal with the vulnerability; On the other hand, strengthen the security mechanism that protect the security services, and identify possible risk sources and risk communication path according to the access associated tree in the analysis process of indirect threat, and then make appropriate protective measures according to risk sources and risk communication path.

## 4 Simulation Experiment

In this paper, we take low-orbit Iridium satellite communications network as the example,  $N_1 \sim N_4$  are satellite nodes,  $N_5$  is a gateway station,  $N_6$  is the user station,  $N_7$  is the system control station, the topology of the network is shown in Fig.4.1,  $N_1$  communicates with  $N_6$  through the gateway station  $N_5$ ,  $N_5$  establish communication with  $N_1$ , take  $N_2$  as the relay route, and then to  $N_4$ , at last, the communication is between  $N_4$  and  $N_6$ ,  $N_7$  plays the role of communication control and management.

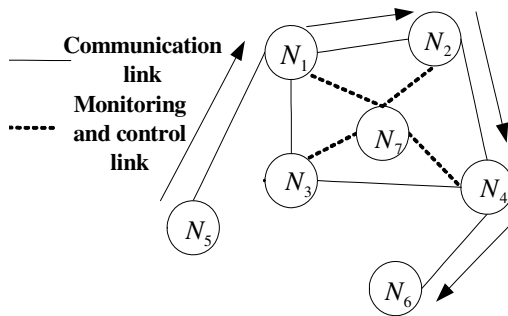


Fig. 4.1. Network topology

Node Ni provides security services SAj, the relative weight wsij service and the harmful levels SAij\_Impact to host after the destruction of service is shown in Tab.4.1, the relative weights of the service obtained from node function using AHP.

**Table 4.1.** security services weight of the node and the harm to the host after the services destruction

Iridium Satellite N <sub>1</sub> ~N <sub>4</sub>			Earth Station N <sub>5</sub> , N <sub>6</sub>			System Control Station N <sub>7</sub>		
SA <sub>j</sub>	ws <sub>ij</sub>	SA <sub>ij</sub> _Impact	SA <sub>j</sub>	ws <sub>ij</sub>	SA <sub>ij</sub> _Impact	SA <sub>j</sub>	ws <sub>ij</sub>	SA <sub>ij</sub> _Impact
Availability	0.26	3	Availability	0.24	3	Availability	0.18	2
Integrity	0.25	3	Integrity	0.23	2	Integrity	0.2	2
Confidentiality	0.2	2	Confidentiality	0.16	1	Confidentiality	0.22	3
Certification	0.15	1	Certification	0.19	2	Certification	0.21	3
Non-repudiation	0.14	1	Non-repudiation	0.18	2	Non-repudiation	0.19	2

The security mechanisms running on each node as shown in Tab.4.2:

**Table 4.2.** The security mechanisms lists of each node

Iridium Satellite N <sub>1</sub> ~N <sub>4</sub>	Earth Station N <sub>5</sub> , N <sub>6</sub>	System Control Station N <sub>7</sub>
Encryption and key management	Access Control	Encryption and key management
Access Control	Authentication	Access Control
Authentication	Error Control	Error Control
Wireless Spread Spectrum		

According to the threat situation of node N<sub>1</sub> ~ N<sub>7</sub> that perceived by FW, IPS, AV, SIT, we can get the existential threat of N<sub>1</sub>, N<sub>5</sub>, N<sub>7</sub>, as shown in Tab. 4.3:

**Table 4.3.** Threat of each node that existential at time t

Time t	N <sub>7</sub>			N <sub>5</sub>			N <sub>1</sub>		
	SNMP Threat	HTTP Threat	TELNET Threat	RPC Threat	FTP Threat	HTTP Threat	FTP Threat	TELNET Threat	DNS Threat
FW	V	V	V	V	V	V	V	V	V
IPS	V	V	V	V	V	V	V	V	V
A-V	V	V	V	V	V	V	V		V
SIT	V	V	V	V	V	V	V	V	V

Reference to credibility of the threat occurring that sensors perceived(knowledge base), after the integration of evidence-based theory, we can obtain reliability value that described the occurrence of threats on each node (probability), as shown in Tab.4.4:

**Table 4.4.** The occurrence probability of threat on each node

	SNMP-Thr	RPC-Thr	FTP-Thr	HTTP-Thr	TELNET-Thr	DNS-Thr
Node $N_7$	0.784			0.857	0.7	
Node $N_5$		0.778	0.692	0.857		
Node $N_1$			0.692		0.7	0.903

Combining the occurrence probability of threat on node  $N_7$  in Tab.7 and the threat that do harm to each security service, we can get the threat occurrence probability of each security service and risk values on  $N_7$ , as shown in Tab.4.5:

**Table 4.5.** Risk of each security service on node  $N_7$

Node $N_7$	Direct threat probability	Indirect threat probability	Threat occurrence probability	$SA_{7,j}$ -Impact	Safeguarding Role Guard $_{7,j}$	Total risk value $RS_{7,j}$
Availability	0.9907	0	0.9907	2	2	0.663798
Integrity	0.9352	0	0.9352	2	3	0.417723
Confidentiality	0.9907	0	0.9907	3	3	0.663798
Certification	0.9907	0	0.9907	3	2	0.995697
Non-repudiation	0.9352	0	0.9352	2	1	1.253168

**Table 4.6.** Risk of each security service on node  $N_5$

Node $N_5$	Direct threat probability	Indirect threat probability	Threat occurrence probability	$SA_{5,j}$ -Impact	Safeguarding Role Guard $_{5,j}$	Total risk value $RS_{5,j}$
Availability	0.9902	0	0.9902	3	2	0.995183
Integrity	0.7778	0	0.7778	2	3	0.347407
Confidentiality	0.9902	0	0.9902	1	1	0.663455
Certification	0.9316	0	0.9316	2	3	0.416125
Non-repudiation	0	0	0			0

**Table 4.7.** Risk of each security service on node  $N_1$

Node $N_1$	Direct threat probability	Indirect threat probability	Threat occurrence probability	$SA_{1,j}$ -Impact	Safeguarding Role Guard $_{1,j}$	Total risk value $RS_{1,j}$
Availability	0.991	0.952	1.000	3	3	0.670
Integrity	0.700		0.700	3	2	0.704
Confidentiality	0.908		0.908	2	3	0.405
Certification	0.908		0.908	1	3	0.203
Non-repudiation	0.700		0.700	1	2	0.235

Similarly, we can get risks of each security service on node  $N_5$  and node  $N_1$ , as shown in Tab.4.6, 4.7:

Therefore, we can get the risk value of network nodes, risk probability, and the risk value and risk probability of the entire network, as shown in Tab.4.8:

**Table 4.8.** Risk value and risk probability of the entire network

	Node $N_1$	Node $N_5$	Node $N_7$	The entire network
Risk $RN_i$	0.494	0.5040	0.7963	0.3021
Risk probability $PN_i$	0.99999967150975	0.9999985502144	0.999999996668952	1

According to Tab.4.8, we can know that the risk value of the entire network is 0.3021, while:

The risk of node  $N_7$  is the maximum,  $N_5$  is secondary to  $N_7$ , the risk of  $N_1$  is minimal. Take node  $N_1$  as the example, we should strengthen the protection of two security services: "integrity" and "availability". For the "integrity", on one hand, we need to add security mechanism of "integrity testing", and on the other hand, fix bugs for the vulnerabilities of "TELNET" service; For the "usability", in addition to direct threats, but also indirect threats. According to the access associated tree methods, we can get the risk source (TELNET and SNMP threats of  $N_7$ , and RPC threat of  $N_5$ ) and risk communication path ( $e(N_7, N_1)$ ,  $e(N_5, N_1)$ ). On the one hand, patch vulnerabilities for the vulnerabilities of service "FTP, TELNET and DNS" on node  $N_1$ , on the other hand, patch vulnerabilities for the vulnerabilities of service "TELNET and SNMP" on node  $N_7$  and service "RPC" on node  $N_5$ . (2) The risk probabilities of node  $N_1$ ,  $N_5$  and  $N_7$  are high, in which  $N_7$  is the highest, need protection.

## 5 Conclusion

This paper based on Network Security Situation Awareness framework, combined with DS Theory and CORAS, proposes a Network Security Risk Assessment Model to identify the threat situation, and do layer by layer risk quantization based on probabilistic risk analysis, introduced the access associated tree of nodes, obtained the source and propagation path of risk, and introduced updated and feedback mechanism of knowledge base to increase the accuracy of risk assessment. The model and method have the following characteristics: (1) get the vulnerability of node according to the analysis on direct threat, and develop measures to fix flaws In a targeted manner;(2) According to the threat that each security service suffered, enhance the appropriate security mechanisms In a targeted manner;(3)According to the risk source and risk path which obtained through the analysis of indirect threat, we can develop measures to reduce, or even eliminate the risk sources and risk path. Simulation results verify the validity of assessment models and methods.

**Acknowledgments.** This paper was supported by a grant from the National Natural Science Foundation of China (60903027) and the CAST Innovation Found (CAST200839).

## References

1. Zhang, Y., Tan, X., Xi, H.: A Novel Approach to Network Security Situation Awareness Based on Multi-perspective Analysis. In: IEEE 2007 International Conference on Computational Intelligence and Security (2007)
2. Liu, X., Wang, H., Lai, J., Liang, Y.: Network Security Situation Awareness Model Based on Heterogeneous Multi-sensor Data Fusion, pp. 1545–1548. IEEE (2007)
3. Lai, J., Wang, H., Zhu, L.: Study of Network Security Situation Awareness Model Based on Simple Additive Weight and Gray Theory, pp. 1545–1548. IEEE (2006)
4. Vraalsen, F., Braber, F.D., Hogganvik, I., Lund, M.S., StøLen, K.: The CORAS Tool-Supported Methodology for UML-Based Security Analysis. SINTEFICT Cooperative and Trusted Systems (February 1992)
5. Sun, L.: An Information Systems Security Risk Assessment Model under Dempster-Shafer Theory of Belief Functions, pp. 109–142 (Spring 2006)
6. Chen, X.Z., Zheng, Q.H., Guan, X.H., et al.: Quantitative hierarchical threat evaluation model for network security. *Journal of Software* 17(4)

# Unit Transient Based Protection for High Voltage Wind Power Transmission Lines

Mingchao Xia<sup>1</sup>, Yong Lin<sup>2</sup>, and Edward W.C. Lo<sup>3</sup>

<sup>1</sup> Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> Co-Author, Hong Kong Polytechnic University, Hong Kong, China

<sup>3</sup> Hong Kong Polytechnic University, Hong Kong, China

**Abstract.** Based on the wavelet transform results of the transient signals from all ends of the line, a unit protection for high voltage power transmission line is presented. The attenuation effects of high voltage transmission line and busbar capacitance to high-frequency current signals are analyzed. The Transient Current Content Ratio (TCCR), Transient Voltage Content Ratio (TVCR) and Transient Voltage Singularity Ratio (TVSR) are deduced according to the analysis of the differences during high frequency voltage transients at each end of the transmission line when in-zone faulted and out-zone faulted. Protection criterion using TCCR and protection criterion using TVSR are proposed, and the principles of protection settings are given. The wavelet modulus maxima of transient signals are used to calculate TCCR and TVSR. Simulation experiments of different faults with EMTP are carried out to test and analyze the protection criteria. A 100kHz sampling rate is used in the simulation experiments. According to the simulation results, the protection criteria prove to be simple, reliable and good for whole line protection.

**Keywords:** transient based protection, current transient, voltage transient, wavelet transform.

## 1 Introduction

Due to the growing capacity of large scale wind farms, for example, more than 500MW wind power generation has been developed in southwest Minnesota, and the fact that wind farms are most commonly located in remote areas and/or at the ends of long sub transmission, transmission, or EHV transmission lines, more high voltage transmission lines are needed to increase the transmission outlet capacity. In Europe, high voltage AC transmission has begun to be used while HVDC is not a feasible solution for the scale of offshore wind farms today. Because of the dynamic character of wind farms, voltage fluctuation and flicker caused by wind power generation could impose impacts on the high voltage transmission lines and may lead to more transient disturbances.

High-voltage transmission lines, with significant characteristics of distributed parameters, will generate abundant high frequency voltage transients and current transients when faults occur. These transient signals contain extensive information



about the faults such as fault type, location, direction and lasting time, etc. In recent years, Transient Based Protections (TBP)[1] - protections using these transients - are widely studied[2,3,4]. TBPs detect the fault-generated voltage and current transient signals to identify faults, and have features of fast response, high accuracy, and immunity to such power-frequency phenomena as fault resistance, power swing, and TA saturation.

Differences in contents of high frequency current signals in time of in-zone fault and out-zone fault have been used to identify faults. However, TBP using single end current transients does not have enough sensitivity or is hard to set under some circumstances such as faults on the end of the line or on the busbar[5]. This will lower the performance of the protection. With the popularization of Fiber Optical communication network in power system, protections using transient signals of all ends of the line become feasible. Unit Transient Based Protection criterion using current transients or voltage transients for two-end and T-type high voltage transmission line is presented. The criterion uses a sampling rate of 100kHz, which will make the criterion easier to be implemented.

## 2 Protection Criteria

Fault components of the air-module of voltage transients are used in the analysis of this paper. Fault components are derived from the signals after fault, and caused merely by fault. The relations between fault voltage components and fault current components detected by protection are determined by the impedance from protection installation point to the fault point, and immune to system voltage and transition resistance of fault. The air-module is the results of module transform of phase signals, and used to eliminate coupling effects between phase signals. The relation between air-module voltage and current signals is the same as the propagation formula of single transmission line[6].

Fault-generated high frequency transient signals will attenuate when propagating along the transmission line and the attenuation increases with the increase in frequency[7]. The transmission parameter of line is:

$$A(\omega) = \exp[-\gamma l] = \exp[-\alpha l] \cdot \exp[-j\beta l] \quad (1)$$

where  $l$  is the distance of propagation,  $\alpha$  is amplitude attenuation parameter called attenuation constant, and  $\beta$  is called phase constant.  $\alpha$  and  $\beta$  are frequency dependent. According to the attenuation of high frequency signals, the fault component current detected by protection is:

$$i(t) = \sum_n i_n(t) = \sum_n I_n \exp[-\alpha_n l] \sin(\omega_n t + \theta_n) \quad (2)$$

where  $I_n$  is the amplitude of current signal at frequency  $\omega_n$  at fault point,  $\alpha_n$  is the attenuation constant of signals at frequency  $\omega_n$ , and  $l$  is the distance from protection installation point to fault point.

At high frequencies, the capacitance and capacitive coupling of high voltage busbars become the dominant factor in the busbar impedance to earth, and a

significant portion of the high frequency current transients will be shunted to earth when passing through the busbar. As a result, this will cause the attenuation of voltage transient signals. The shunt effect of busbar capacitance to high frequency current is show in Fig.1, where  $K_i$  is the ratio of outgoing current through busbar to incoming current[5].

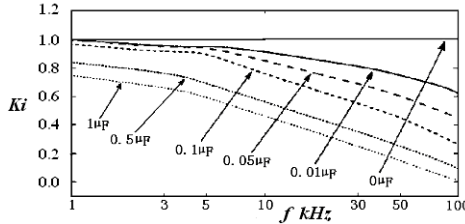


Fig. 1. Attenuation of busbar capacitance to the incoming current

**A. Analysis Of Two-End Transmission Line Protection**

A simplified model of two-end high voltage transmission line is shown in Fig.2, where MN is the line section for protection, and currents through breakers 1, 2 and 3 are  $i_M$ ,  $i_N$ , and  $i_3$  respectively (positive when current is from busbar to line).  $G_1$  and  $G_2$  are two power sources, and  $C_Z$  is the stray busbar capacitance. F1 is in-zone fault point and F2 is out-zone fault point.

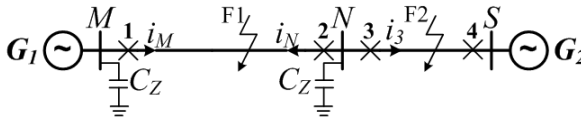


Fig. 2. Simplified model of two-end high voltage transmission line

**A1. Analysis of In-zone Fault**

During in-zone faults, fault signal components are shown in Fig.3.

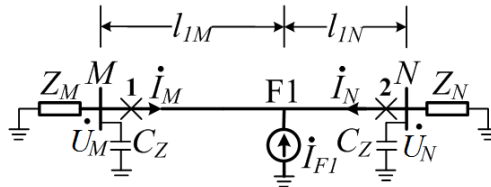


Fig. 3. Fault signal components of in-zone fault of two-end line

Where distances from fault point  $F_1$  to end M and to end N are  $l_{IM}$  and  $l_{IN}$ . The fault current components through protections 1 and 2 at frequency  $\omega$  are  $I_{M\omega}$  and  $I_{N\omega}$ , and fault voltage components at busbars M and N at frequency  $\omega$  are  $U_{M\omega}$  and  $U_{N\omega}$

respectively.  $Z_M$  and  $Z_N$  are the equivalent impedances of busbars M and N to earth (without stray capacitance  $C_z$ ).

Assume that the current branch factor at fault point  $F_1$  to line  $F_1$ -M section is  $K_{M1}$ , the current content of fault current source  $I_{F1}$  at frequency  $\omega$  is  $I_{F1}A_\omega$ , and the amplitude attenuation parameter of line MN section at frequency  $\omega$  is  $\alpha_\omega$ . Then we get:

$$I_{M\omega} = I_{F1}A_\omega K_{M1} \exp[-\alpha_\omega l_{1M}] \tag{3}$$

Similarly, assuming the current branch factor at fault point  $F_1$  to line  $F_1$ -N section is  $K_{N1}$ , we get:

$$I_{N\omega} = I_{F1}A_\omega K_{N1} \exp[-\alpha_\omega l_{1N}] \tag{4}$$

As we can see, the fault components of current transients detected by the protections at the two ends of the line attenuate along with the frequency. We define the transient current content ratio (TCCR) as the ratio of fault component of current transients at different high frequencies  $\omega_L$  and  $\omega_H$ :

$$KI_{LH} = I_{\omega_L} / I_{\omega_H} \tag{5}$$

At high frequency, the impedance of transmission line is mainly inductance, so the current branch factor of transmission line at frequencies  $\omega_L$  and  $\omega_H$  is approximately the same, that is,  $K_{M1H} \approx K_{M1L}$ ,  $K_{N1H} \approx K_{N1L}$ . So during in-zone faults, the TCCRs are:

$$\begin{cases} KI_{LHM-in} = I_{M\omega_L} / I_{M\omega_H} = A_{\omega_L} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{1M}] / A_{\omega_H} \\ KI_{LHN-in} = I_{N\omega_L} / I_{N\omega_H} = A_{\omega_L} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{1N}] / A_{\omega_H} \end{cases} \tag{6}$$

where  $A_{\omega_L}$  and  $A_{\omega_H}$  are the current contents of fault current source  $I_{F1}$  at frequencies  $\omega_L$  and  $\omega_H$ , and they depend on the fault.  $\alpha_{\omega_L}$  and  $\alpha_{\omega_H}$  are the amplitude attenuation parameters of line MN section at the two frequencies, and as we know,  $\alpha_{\omega_L} < \alpha_{\omega_H}$ .

When using a high frequency  $\omega$ ,  $|Z_M| \gg 1/(j\omega C_z)$ , the fault voltage component on busbar M at frequency  $\omega$  is approximately:

$$\dot{U}_{M\omega} = -\dot{I}_{M\omega} / (j\omega C_z) \tag{7}$$

And we get:  $U_{M\omega} = I_{F1}A_\omega K_{M1} e^{-\alpha_\omega l_{1M}} / (\omega C_z)$  (8)

Similarly, we get:  $U_{N\omega} = I_{F1}A_\omega K_{N1} e^{-\alpha_\omega l_{1N}} / (\omega C_z)$  (9)

Same to TCCR, we define the transient voltage content ratio (TVCR) as the ratio of fault component of voltage transients at different high frequencies  $\omega_L$  and  $\omega_H$ :

$$KU_{LH} = U_{\omega_L} / U_{\omega_H} \tag{10}$$

So, when in-zone faulted, the TVCRs are:

$$\begin{cases} KU_{LHM-in} = \frac{A_{\omega_L} \omega_H}{A_{\omega_H} \omega_L} e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{1M}} \\ KU_{LHN-in} = \frac{A_{\omega_L} \omega_H}{A_{\omega_H} \omega_L} e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{1N}} \end{cases} \quad (11)$$

## A2. Analysis of Out-Zone Fault

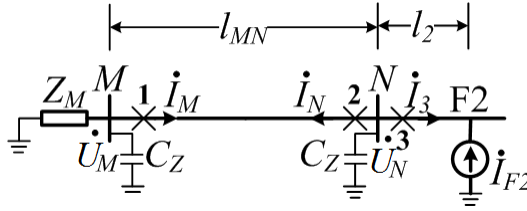


Fig. 4. Fault signal components of an out-zone fault of two-end line

During out-zone faults, fault signal components are shown in Fig.4; distance from fault point  $F_2$  to end N is  $l_2$ , and the length of line MN section is  $l_{MN}$ . At frequency  $\omega$ , the fault current component through protection 3 is  $I_{3\omega}$ , and the fault current component through protection 2 is  $I_{N\omega}$ . The equivalent impedance of line MN section at frequency  $\omega$  is  $Z=R+j\omega L$ .

When using a high frequency,  $|Z_M| \gg 1/(j\omega C_z)$ , and considering the shunt effect of busbar capacitance, approximately, we get:

$$\begin{aligned} \dot{I}_{N\omega} &= -\dot{I}_{3\omega} \frac{(1/j\omega C_z)}{(j\omega L - 1/j\omega C_z - 1/j\omega C_z)} \\ &= j\dot{I}_{3\omega}/(\omega^2 LC_z - 2) \end{aligned} \quad (12)$$

Then, at protection 1 of end M, in consideration of the attenuation effect of line MN section, fault current component and fault voltage component are:

$$I_{M\omega} = I_{N\omega} \exp[-\alpha_{\omega} l_{MN}] = \frac{I_{3\omega}}{\omega^2 LC_z} \exp[-\alpha_{\omega} l_{MN}] \quad (13)$$

Similarly to (3), considering the attenuation effect of line NS section, we get:

$$I_{3\omega} = I_{F_2} A_{2\omega} K_{N_2} \exp[-\alpha_{2\omega} l_2] \quad (14)$$

where,  $K_{N_2}$  is the current branch factor at fault point  $F_2$  to line  $F_2$ -N section,  $I_{F_2} A_{2\omega}$  is the current content of fault current source  $I_{F_2}$  at frequency  $\omega$ , and  $\alpha_{2\omega}$  is the amplitude

attenuation parameter of line NS section at frequency  $\omega$  So we get the TCCR at end M and end N during out-zone faults:

$$\begin{cases} KI_{LHM-out} = \frac{A_{2\omega_L}}{A_{2\omega_H}} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}] \frac{\omega_H^2}{\omega_L^2} \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2] \\ KI_{LHN-out} = \frac{A_{2\omega_L}}{A_{2\omega_H}} \frac{\omega_H^2}{\omega_L^2} \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2] \end{cases} \quad (15)$$

where,  $A_{2\omega_L}$  and  $A_{2\omega_H}$ , which depend on the fault, are the current contents of fault current source  $I_{F2}$  at frequencies  $\omega_L$  and  $\omega_H$ .  $\alpha_{2\omega_L}$  and  $\alpha_{2\omega_H}$ , which depend on the line parameter, are the amplitude attenuation parameters of line NS section at the two frequencies, as we know,  $\alpha_{2\omega_L} < \alpha_{2\omega_H}$ .

When using a high frequency,  $|Z_M| \gg 1/(j\omega C_z)$ ,  $\omega L \gg R$ ,  $\omega L \gg 1/(j\omega C_z)$ . The fault voltage component at busbar N is:

$$\dot{U}_{N\omega} \approx -\dot{I}_{3\omega} / (j\omega C_z) \quad (16)$$

that is, 
$$U_{N\omega} \approx I_{3\omega} / (\omega C_z) \quad (17)$$

According to (13) and (14), we get

$$U_{N\omega} = I_{F2} A_{2\omega} K_{N2} \exp[-\alpha_{2\omega} l_2] / (\omega C_z) \quad (18)$$

$$U_{M\omega} = \frac{I_{F2} A_{2\omega} K_{N2} e^{-\alpha_{2\omega} l_2} e^{-\alpha_{\omega} l_{MN}}}{\omega C_z} \bullet \frac{1}{\omega^2 LC_z} \quad (19)$$

And we can get the transient voltage content ratios at end M and end N during out-zone faults as:

$$\begin{cases} KU_{LHM-out} = \frac{A_{2\omega_L}}{A_{2\omega_H}} \frac{\omega_H}{\omega_L} e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}} \frac{\omega_H^2}{\omega_L^2} e^{(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2} \\ KU_{LHN-out} = \frac{A_{2\omega_L}}{A_{2\omega_H}} \frac{\omega_H}{\omega_L} e^{(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2} \end{cases} \quad (20)$$

### A3. Protection Criterion for Two-End Line

According to (6) and (15) we can see, using current transients,  $KI_{LHN}$  of out-zone fault (as in (15)) is  $\omega_H^2 / \omega_L^2$  times to  $KI_{LHN}$  of in-zone fault (as in (6)) during in-zone faults and out-zone fault occur at same distance to end N. We can use  $KI_{LHN}$  to achieve protection: small  $KI_{LHN}$  means in-zone fault while big  $KI_{LHN}$  means out-zone fault. But under some circumstances, such as in-zone fault occurring far from busbar and out-zone fault occurring near busbar,  $KI_{LHN}$  of out-zone fault may be smaller than  $KI_{LHN}$  of in-zone fault:

when in-zone fault point F1 far from end N,  $l_{1N} \approx l_{MN}$ :

$$KI_{LHN-in} = A_{\omega_L} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}] / A_{\omega_H} \quad (21)$$

when out-zone fault point F2 near end N,  $l_2 \approx 0$ :

$$KI_{LHN-out} = A_{\omega_L} \omega_H^2 / (A_{\omega_H} \omega_L^2) \quad (22)$$

Some line parameters may cause  $\exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}] > \omega_H^2 / \omega_L^2$ , and the protection at end N using  $KI_{LHN}$  cannot operate properly. We use the TCCR of the opposite end to aid the protection judgment: TCCRs of end N and end M are both small during in-zone faults, TCCRs of end N and end M are both big during out-zone faults, and TCCR of end M (far from fault point) is much larger. So the protection criterion is:

$$KI_{LHM} < KI_{set} \text{ and } KI_{LHN} < KI_{set} \quad (23)$$

In this criterion, when out-zone fault point F2 is near end N,  $l_2 \approx 0$ ,  $KI_{LHN}$  may be small, but  $KI_{LHM}$  is much larger:

$$KI_{LHM-out} = A_{2\omega_L} \omega_H^2 \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_{MN}] / (A_{2\omega_H} \omega_L^2) \quad (24)$$

So we get:

$$\begin{aligned} A_{\omega_L} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}] / A_{\omega_H} &< KI_{set} \quad \text{and} \\ KI_{set} &< \omega_H^2 A_{\omega_L} \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}] / (\omega_L^2 A_{\omega_H}) \end{aligned} \quad (25)$$

Alternative protection criterion is using voltage transients. Comparing (11) with (20), we can see that, when same faults occur at in-zone point F<sub>1</sub> and out-zone point F<sub>2</sub>, the voltage contents at frequencies  $\omega_L$  and  $\omega_H$  of the transient voltage signals detected at busbar near fault point (*i.e.*, busbar N) are very similar, and the transient voltage content ratios  $KU_{LH}$  at this busbar during in-zone fault and out-zone fault differ little. But at busbar far from the out-zone fault point (*i.e.*, busbar M), the transient voltage content ratio  $KU_{LH}$  during out-zone fault is much bigger than that during in-zone fault. That is, the transient voltage content ratios  $KU_{LHM}$  and  $KU_{LHN}$  differ little during in-zone fault, but they differ greatly during out-zone fault. We define the transient voltage singularity ratio (TVSR) as:  $\lambda = \max(KU_{LHM}, KU_{LHN}) / \min(KU_{LHM}, KU_{LHN})$  (26)

where  $\max(KU_{LHM}, KU_{LHN})$  and  $\min(KU_{LHM}, KU_{LHN})$  are the maximum and minimum of  $KU_{LHM}$  and  $KU_{LHN}$ . According to the analysis above, we get:

$$\begin{cases} \text{in-zone faulted} : \lambda = \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN} - l_{1N}] \\ \text{out-zone faulted} : \lambda = \exp[(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN} \omega_H^2 / \omega_L^2] \end{cases} \quad (27)$$

$\lambda > 1$ , it shows the difference of the transient voltage singularities detected at the two ends of the line. During in-zone fault,  $\lambda$  is smaller, whose maximum value is  $e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}}$ ; during out-zone fault,  $\lambda$  is bigger, and changes little. The value of  $\lambda$  is influenced by  $\alpha_{\omega_H}$ ,  $\alpha_{\omega_L}$  and fault location, and denotes the characteristic of the transmission line.

Using the transient voltage singularity ratio  $\lambda$ , we get the protection criterion:

$$\lambda < \lambda_{set} \quad (28)$$

And we get the range of the protection setting  $\lambda_{set}$ :

$$e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}} < \lambda_{set} < e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}} \frac{\omega_H^2}{\omega_L^2} \tag{29}$$

**B. Analysis of T-Type Transmission Line Protection**

Simplified model of T-type high voltage transmission line is shown in Fig.5, where MNR is the line section of protection, and currents through breakers 1, 2, 3, and 4 are  $i_M, i_N, i_R$  and  $i_4$  (positive when current is from busbar to line).  $G_1, G_2$ , and  $G_3$  are three power sources, and  $C_Z$  is the stray busbar capacitance. F1 is in-zone fault point and F2 is out-zone fault point.

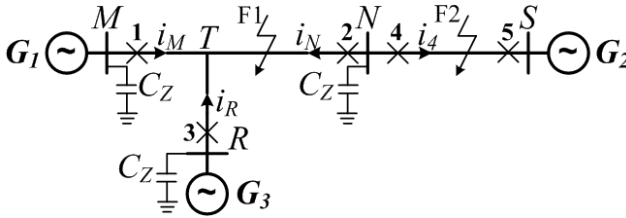


Fig. 5. Simplified model of T-type high voltage transmission line

**B1. Analysis of In-Zone Fault**

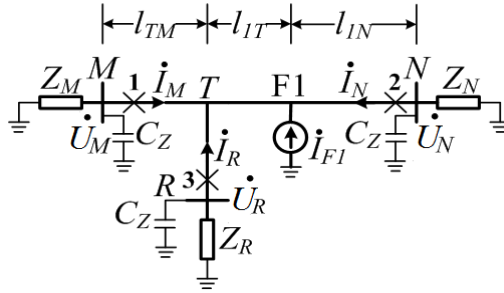


Fig. 6. Fault signal components of in-zone fault of T-type line

In Consideration of fault point inside MNR section, such as F1 in NT section, the fault components are shown in Fig.6. Distances from fault point F1 to point T and to end N are  $l_{IT}$  and  $l_{IN}$ . The fault current component through protection 1 at frequency  $\omega$  is  $I_{M\omega}$ .  $Z_{M\omega} = R_{\omega} + j\omega L_{\omega}$  is the equivalent impedance of busbar M to earth (without stray capacitance). Assuming that the current branch factor of  $I_{F1}$  at fault point F1 to line F1-T section is  $K_{T1}$ , the current branch factor at point T to line MT section is  $K_{MT}$ , the current content of fault current source  $I_{F1}$  at frequency  $\omega$  is  $I_{F1}A_{\omega}$ , the attenuation constant of line NT section at frequency  $\omega$  is  $\alpha_{\omega_{NT}}$  and the attenuation constant of line MT section at frequency  $\omega$  is  $\alpha_{\omega_{MT}}$ , we get:

$$I_{M\omega} = I_{F1} A_{\omega} K_{T1} \exp[-\alpha_w l_{1T}] K_{MT} \exp[-\alpha_w l_{TM}] \quad (30)$$

Similarly, at end R and end N we get:

$$I_{R\omega} = I_{F1} A_{\omega} K_{T1} \exp[-\alpha_w l_{1T}] K_{RT} \exp[-\alpha_w l_{TR}] \quad (31)$$

$$I_{N\omega} = I_{F1} A_{\omega} K_{N1} \exp[-\alpha_w l_{1N}] \quad (32)$$

Similarly at high frequency, the impedances of busbars M, N and R to earth are very low (about  $1/j\omega C_z$ ), and the impedance of transmission line is mainly inductance. Thus the impedances of transmission line play a crucial role in the current branch factors  $K_{T1}$  and  $K_{MT}$ , and make them approximately unchanged at frequencies  $\omega_L$  and  $\omega_H$ . That is,  $K_{T1H} \approx K_{T1L}$ ,  $K_{MTH} \approx K_{MTL}$ . So we get:

$$\begin{cases} KI_{LHM-in} = I_{M\omega_L} / I_{M\omega_H} = A_{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1M}] / A_{\omega_H} \\ KI_{LHR-in} = I_{R\omega_L} / I_{R\omega_H} = A_{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1R}] / A_{\omega_H} \\ KI_{LHN-in} = I_{N\omega_L} / I_{N\omega_H} = A_{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1N}] / A_{\omega_H} \end{cases} \quad (33)$$

where,  $l_{1M} = l_{1T} + l_{TM}$  and  $l_{1R} = l_{1T} + l_{TR}$ . We can see that (33) is similar to (6).

The voltage signal at busbar N is the same as that in Fig.3. The TVCR at busbar N is the same as (11). For voltage signals at busbar M:

$$U_{M\omega} = I_{F1} A_{\omega} K_{MT} K_{T1} \exp[-\alpha_w l_{1M}] / (\omega C_z) \quad (34)$$

Similar to (30) and (31),  $K_{T1H} \approx K_{T1L}$  and  $K_{MTH} \approx K_{MTL}$ . So we get:

$$\begin{cases} KU_{LHM-in} = \frac{A_{\omega_L}}{A_{\omega_H}} \frac{\omega_H}{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1M}] \\ KU_{LHN-in} = \frac{A_{\omega_L}}{A_{\omega_H}} \frac{\omega_H}{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1N}] \\ KU_{LHR-in} = \frac{A_{\omega_L}}{A_{\omega_H}} \frac{\omega_H}{\omega_L} \exp[(\alpha_{w_H} - \alpha_{w_L}) l_{1R}] \end{cases} \quad (35)$$

## B2. Analysis of Out-Zone Fault

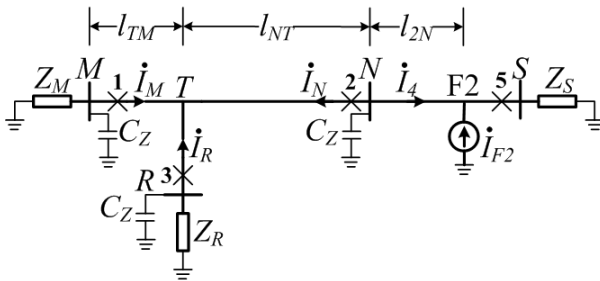


Fig. 7. Fault signal components of out-zone fault of T-type line



In consideration of fault point outside MNR section, such as F2 in NS section, the fault components are shown in Fig.7. Considering the shunt effect of busbar capacitance at bus N, approximately we get:

$$\dot{I}_{N\omega} \approx -\dot{I}_{4\omega}(1/j\omega C_z)/(j\omega L) = j\dot{I}_{4\omega}/(\omega^2 LC_z) \tag{36}$$

Considering the attenuation effect of line NS, we get:

$$I_{4\omega} = I_{F2} A_{2\omega} K_{N2} \exp(-\alpha_{\omega NS} l_{2N}) \tag{37}$$

where,  $A_{2\omega}$  is current content of fault current source  $I_{F2}$  at frequency  $\omega$  and  $K_{N2}$  is the current branch factor at fault point F2 to line F2-N.  $\alpha_{\omega NS}$  is the attenuation constant of line NS.  $I_N$  propagates along line MNR, and reach ends M and R. Similar to (18)-(20), we get:

$$\begin{cases} KI_{LHM-out} = A_{2\omega_L} \omega_H^2 \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_{MN}] \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2] / (A_{2\omega_H} \omega_L^2) \\ KI_{LHR-out} = A_{2\omega_L} \omega_H^2 \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_{RN}] \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2] / (A_{2\omega_H} \omega_L^2) \\ KI_{LHN-out} = A_{2\omega_L} \omega_H^2 \exp[(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2] / (A_{2\omega_H} \omega_L^2) \end{cases} \tag{38}$$

where,  $l_{MN} = l_{MT} + l_{TN}$  and  $l_{RN} = l_{RT} + l_{TN}$ . We can see (38) is similar to (15).

According to the analysis above, for voltage transients, the current branch factor at node T does not affect the TVCR at busbars M and R:

$$\begin{cases} KU_{LHM-out} = \frac{A_{2\omega_L} \omega_H}{A_{2\omega_H} \omega_L} e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{MN}} \frac{\omega_H^2}{\omega_L^2} e^{(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2} \\ KU_{LHN-out} = \frac{A_{2\omega_L} \omega_H}{A_{2\omega_H} \omega_L} e^{(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2} \\ KU_{LHR-out} = \frac{A_{2\omega_L} \omega_H}{A_{2\omega_H} \omega_L} e^{(\alpha_{\omega_H} - \alpha_{\omega_L})l_{RN}} \frac{\omega_H^2}{\omega_L^2} e^{(\alpha_{2\omega_H} - \alpha_{2\omega_L})l_2} \end{cases} \tag{39}$$

**B3. Protection Criterion for T-Type Line**

As (33) is similar to (6) and (38) is similar to (15), we can get the transient current protection criterion for T-type line:

$$KI_{LHM} < KI_{set} \text{ and } KI_{LHN} < KI_{set} \text{ and } KI_{LHR} < KI_{set} \tag{40}$$

where protection setting  $KI_{set}$  is the same as (25). And now the TVSR  $\lambda$  is:

$$\lambda = \max(KU_{LHM}, KU_{LHN}, KU_{LHR}) / \min(KU_{LHM}, KU_{LHN}, KU_{LHR}) \tag{41}$$

The protection criterion is the same as expression (28), and the protection setting range is the same as (29) (assuming  $l_{MN} \geq l_{MR} \geq l_{NR}$ ).

**C. Signal Processing**

Wavelet transform is used to decompose high frequency voltage signals and current signals to different frequency banks. Wavelet transform is a linear transform[8], and the results of wavelet transform under different scales are related to the amplitudes of the signals in the frequency banks according to the scale. So we can get the TVCRs

and TVCRs through the ratio of the wavelet transform results under different scales. We use the ratio of wavelet modulus maxima under scale 3 and scale 1- $WTKU$  and  $WTKI$ - to calculate  $KU_{LH}$  and  $KI_{LH}$ . According to the multi-scale formula of wavelet transform, we get:

$$KU_{LH}=2WTKU \tag{42}$$

And 
$$KI_{LH}=2WTKI \tag{43}$$

In the analysis above, we use fault components of the voltage transients and current transients detected on all ends of the line. As we did not assume that the transients were detected simultaneously, and in consideration of the propagation time of the signal along the transmission line, the protection criterion in this paper does not need critically synchronized sampling on each end of the line.

According to the shunt effect of busbar capacitance, a sampling rate of 100kHz is used in this paper to sample the voltage transients. The frequency banks associated with wavelet transform scale 3 and scale 1 are 6.25kHz~12.5kHz and 25kHz~50kHz. And the protection setting should be:

$$KI_{set} = 4\exp[(\alpha_{wH} - \alpha_{wL})l_{MN}] \tag{44}$$

$$\lambda_{set} = 2e^{(\alpha_{wHNT} - \alpha_{wLNT})l_{NT} + (\alpha_{wHMT} - \alpha_{wLMT})l_{MT}} \tag{45}$$

### 3 Simulation Results

A simplified model (using frequency-dependent line model) of 500kV two-end transmission line shown in Fig.8 is used to simulate the two-end line.

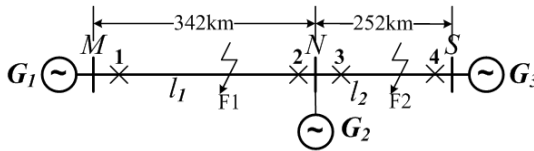


Fig. 8. Two-end Line model for EMTP simulation

Line model and tower configuration in Figure 8 are derived from 500kV Ping-Wu line in China. Section MN is the section of protection.  $G1=1699MVA$ ,  $G2=704MVA$ ,  $G3=178MVA$ ,  $F1$  and  $F2$  are two different fault points, and distance from  $F1$  to end  $M$  is  $l_1$ , from  $F2$  to end  $N$  is  $l_2$ .  $MN$  length is 342km, while  $NS$  length is 252km.

Different fault types, such as single phase to earth, phase-phase, 3-phase, different fault impedances (low and high impedance), and different fault start times are used for simulation experiments with PSCAD/EMTDC with a sampling rate of 100kHz and a 2% white noise. According to line parameter, the protection settings are  $KI_{set}=5.2$  and  $\lambda_{set} = 2.6$ .

Tab.1 shows the Transient Current Protection(TCP) results of phase C-to-earth permanent fault with fault resistance 10 ohms as an example, and Tab.2 shows the Transient Voltage Protection(TVP) results of the phase C-to-earth permanent fault with fault resistance 10 ohms as an example.

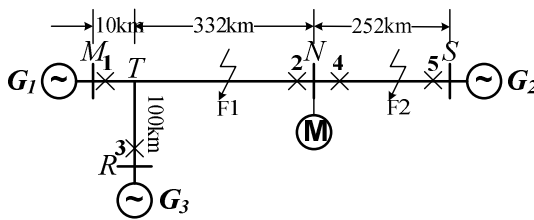
**Table 1.** TCP Judgment Results of phase C fault of two-end line

No.	Content	$KI_{LHM}$	$KI_{LHN}$	Judgement
1.1	$l_1=1\text{km}$	2.370	3.948	in-zone
1.2	$l_1=10\text{km}$	2.482	3.778	in-zone
1.3	$l_1=171\text{km}$	2.296	2.722	in-zone
1.4	$l_1=282\text{km}$	2.940	2.128	in-zone
1.5	$l_1=341\text{km}$	3.222	3.012	in-zone
1.6	$l_2=1\text{km}$	8.194	9.546	in-zone
1.7	$l_2=10\text{km}$	6.780	7.674	out-zone
1.8	$l_2=126\text{km}$	6.500	7.234	out-zone

**Table 2.** TVP Judgment Results of phase C fault of two-end line

No.	Content	$KI_{LHM}$	$KI_{LHN}$	$\lambda$	Judgement
2.1	$l_1=1\text{km}$	9.446	8.964	1.054	in-zone
2.2	$l_1=10\text{km}$	9.814	8.354	1.175	in-zone
2.3	$l_1=171\text{km}$	9.530	7.576	1.258	in-zone
2.4	$l_1=282\text{km}$	9.864	7.843	1.258	in-zone
2.5	$l_1=341\text{km}$	10.244	9.552	1.072	in-zone
2.6	$l_2=1\text{km}$	39.806	9.552	4.167	in-zone
2.7	$l_2=10\text{km}$	34.540	7.686	4.494	out-zone
2.8	$l_2=126\text{km}$	38.268	7.246	5.281	out-zone

A simplified model of 500kV T-type transmission line shown in Fig. 9 is used as simulation model for T-type lines.



**Fig. 9.** Line model for EMTP simulation

where line MNR section is the section for protection.  $G_1=1699\text{MVA}$ ,  $G_3=704\text{MVA}$ ,  $G_4=178\text{MVA}$ , F1 and F2 are two different fault points. MT is 10km long, NT is 332km long, and RT is 100km long. NS is 252km long.

Tab.3 shows the Transient Current Protection(TCP) results of the phase C to earth permanent fault of T-type line with fault resistance 10 ohms as an example, and Table 4 shows the Transient Voltage Protection(TVP) results of the phase C to earth permanent fault of T-type line with fault resistance 10 ohms as an example.

**Table 3.** TCP Judgment Results of phase C fault of T-type line

No.	Content	$KI_{LHR}$	$KI_{LHM}$	$KI_{LHN}$	Judgement
3.1	$l_{N1}=1\text{km}$	2.365	2.388	3.892	in-zone
3.2	$l_{N1}=10\text{km}$	2.446	2.524	3.647	in-zone
3.3	$l_{N1}=322\text{km}$	2.711	2.676	2.972	in-zone
3.4	$l_{M1}=331\text{km}$	2.983	3.129	3.101	in-zone
3.5	$l_{M1}=5\text{km}$	3.134	2.986	3.293	in-zone
3.6	$l_{R1}=10\text{km}$	3.202	3.109	3.192	in-zone
3.7	$l_{R1}=90\text{km}$	3.065	3.204	3.309	in-zone
3.8	$l_2=1\text{km}$	8.326	8.205	9.644	in-zone
3.9	$l_2=10\text{km}$	7.033	6.892	7.593	out-zone
3.10	$l_2=60\text{km}$	6.342	6.102	6.812	out-zone
3.11	$l_2=126\text{km}$	6.982	6.621	7.238	out-zone

**Table 4.** TVP Judgment Results of phase C fault of T-type line

No.	Content	$KI_{LHR}$	$KI_{LHM}$	$KI_{LHN}$	$\lambda$	Judgement
4.1	$l_{M1}=5\text{km}$	11.852	9.424	13.572	1.44	in-zone
4.2	$l_{R1}=5\text{km}$	9.368	10.823	13.583	1.45	in-zone
4.3	$l_{R1}=95\text{km}$	10.211	11.536	14.968	1.47	in-zone
4.4	$l_{N1}=5\text{km}$	14.755	14.028	10.546	1.40	in-zone
4.5	$l_{N1}=327\text{km}$	11.875	10.778	11.243	1.06	in-zone
4.6	$l_{N2}=5\text{km}$	44.581	39.986	10.534	4.23	out-zone
4.7	$l_{N2}=126\text{km}$	48.385	41.943	9.087	5.32	out-zone

The simulation and protection judgment results prove that the protection criterion presented in this paper can be used to judge properly under different fault locations, fault types and system status, and can to protect the whole line.

## 4 Conclusions

The voltage transients and current transients from all ends of the high voltage transmission line are analyzed and used to get the TVSR and TCCR, which are used to form unit transient voltage protection and transient current protection. The novel unit transient based protection criteria presented in this paper are suitable for the protection of two-end line and T-type line, and do not need critically synchronized sampling at each end. Simulation results of EMTP program prove that the criteria can be used to determine in-zone faults and out-zone faults properly. For different fault

instances, the criteria can be used to determine properly and to protect the whole line. The criteria operate rapidly and reliably with good selectivity.

**Acknowledgments.** Corresponding author: Xia Mingchao (mchxia@bjtu.edu.cn).  
Co-corresponding author: Lin Yong(yonglin@yeah.net).

## References

1. Bo, Z.Q., Weller, G., Dai, F.T., et al.: Transient Based Protection for Transmission Lines. In: POWERCON 1998, Beijing, China, pp. 1067–1071 (1998)
2. Fu, L., He, Z., Bo, Z.: Novel Approach to Fault Classification in EHV Transmission Line based on Multi-information Measurements of Fault Transients. In: APPEEC 2009, Asia-Pacific, Wuhan, China, March 28-30, pp. 1–4 (2009)
3. Xia, M.C., Huang, Y.Z.: A new protection criterion for High-Voltage transmission line using two-end voltage transients. In: DPSP 2004, pp. 84–87 (2004)
4. Solanki, M., Song, Y.H.: Transient protection of EHV transmission line using discrete wavelet analysis. In: IEEE Power Engineering Society General Meeting, Toronto, Canada, July 13-17, pp. 1868–1873 (2003)
5. Bo, Z.Q.: A new non-communication protection technique for transmission lines. IEEE Trans. on Power Delivery 3(4), 1073–1078 (1998)
6. Marti, J.: Accurate Modeling of Frequency Dependent Transmission Lines in Electromagnetic Transients Simulation. IEEE Trans. on Power Apparatus and Systems PAS-101(1), 147–155 (1982)
7. Gustavsen, B., Irwin, G., Mangelr, R., et al.: Transmission Line Models for the Simulation of Interaction Phenomena Between Parallel AC and DC Overhead Lines. In: Proceedings of IPST 1999, pp. 61–67 (1999)
8. Mallat, S.: A wavelet tour of signal processing. Academic Press, San Diego (1998)

# Assessment of EXPO's Impact on Regional Economy

Shiyun Yu<sup>1</sup> and Hao Li<sup>2</sup>

<sup>1</sup> Jingsiyuan dormitory building, 607, Shanghai University of Finance & Economics,  
777 GuoDing Road, YangPu District,  
Shanghai Province, China

<sup>2</sup> North 4 dormitory building, 238, Shanghai University of Finance & Economics,  
777 GuoDing Road, YangPu District,  
Shanghai Province, China

**Abstract.** The Expo is not only an festival about science, culture and history, but also an pivot economic event. The Expo works as a catalyst to the endemic economics perennially and profoundly as it will boost tourism, promote tertiary industry, attract investments and so on. Based on the statistical data about the beginning and mid-term, this paper use Grey Model, Entropy Method, Comprehensive quantitative analysis to simulate, predict and test the ending period data by computer softwares which make this model integrate.

**Keywords:** Expo impact, Comprehensive quantitative analysis, EntropyMethod, GreyModel, Expo-cycle efficiency.

## 1 Introduction

In the past, the influence of major economic events are commonly estimated by expenditure which directly transform the investments in the project through the multiplier effect to the national economy at all levels. But since the Shanghai Expo is so big and profound an event, and it has not end its effect to the three major industries in Shanghai, we should not estimate its effects normally.

In this way, this paper use an indirect method to estimate the Expo effects by comparison between the economic development in the economic cycle and economic development before the cycle.

Expo will boost the economy in a long-term, phased process, so the assessment should not only consider the data in the year that Expo was held, but to look at a complete economic cycle. By accessing to the results found by previous research, the former Expo period is usually divided into pre-Expo stage, the Expo stage and after-stage of the Expo. The Expo will impact each expo stage in different ways. So we estimate the effects by phases to be strict and precise.

## 2 Nomenclature

$X''$  : original target value

$\Delta X$  : Expo incremental

$X$  : values of economic indicators

- $X'$  : values of economic indicators after removing inflation
- $\tilde{X}$  : forecasted economic Indicators
- MPC : Marginal consumer price index
- $\Delta G$  : Government purchase
- $\Delta Y$  : increase on consumption
- $f_{ij}$  raw data of decision matrix, indicator j in year i
- $d_{ij}$  : standard data of decision matrix, the standardization of indicator j in year i
- $E_j$  : the entropy of the j-information value indicators.
- $w_j$  : the j value of indicators of empowerment
- $v_i$  : the i-year comprehensive evaluation index.

### 3 Basic Assumptions

- 1) Assuming Expo is the only major event taking place in Shanghai during the Expo period,
- 2) Assuming that if the World Expo did not take place, regional economic indicators will increase under the original speed.
- 3) Assuming the level of consumption is associated with quality of life.

#### 3. Evaluation model of Shanghai World Expo

##### 1. Select economical indicators

Indicators used to describe regional economic are industry output, employment and the level of foreign trade and investment. Especially industry is divided into primary, secondary and tertiary industries.

This paper conduct quantitative analysis of these economic indicators to assess the influence of the Expo.

##### 2. Removal inflation factors.

$$X' = \frac{X''}{CPI}$$

##### 3. Remove random effects

Random events occurs in Socio-economic system. In order to eliminate the random effect on our analysis, do 3 times moving average of the data for smoothing.

$$X_{ij} = \frac{\sum_{n=i-3}^{n=i} X_{nj}}{3}$$

$X_{ij}$  values of economic indicators after removing random effects

#### 4 Index System and Analysis Phases of the Economic Cycle Expo

Construct Expo incremental  $\Delta X$  and Expo incremental strength index S,

$$\Delta X = X - \tilde{X}$$

$\Delta X$	expo incremental
$X$	values of economic indicators
$\tilde{X}$	forecasted economic Indicators

The expo incremental is an absolute indicator. However, according to the economies of scale, the development level in different places are different, Expo incremental does not have direct economic inter-regional comparability. Therefore, we build Expo incremental strength index.

Expo incremental strength index=increment caused by the Expo/original size of the regional economy.:

$$S_{ij} = \frac{\Delta X_i}{X} = \frac{X - \tilde{X}}{X}$$

#### Analysis of the Economic Cycle Expo

1) The pre-Expo stage indicates the preparatory period which usually starts about six years before the opening. At this stage the building of Expo venues and related facilities have a strong influence on the amount of investment.

2) The Expo stage is the year Expo is held. In this period economics growth is mainly driven by the consumption, such as tourism, trade, telecommunications and other aspects of consumptions

3) The after-stage of the Expo is 2 years or even longer period of time after the Expo, Constrained by the lack of demand, the host city may have idle real estate, tourism downturn and other low effect. However, with proper policy method and appropriate measures to deal with sustainable development issues, these problems can be avoided.

#### 5 The Pre-expo Stage

Calculate the expo incremental

##### 1. Grey Relational Analysis

For a more reasonable assessment of Shanghai's economic development, converted investment in fixed assets and other categories of indicators into consumption growth caused by investment value, calculated by the multiplier effect from the perspective of input-output analysis of investment in Shanghai's economy impact.



It is clear that total fixed asset investment value of the multiplier effect can represent the Output effect of investment made in the pre-expo stage. Thus we use the index as a reference series.

Calculate correlation of the indicators and sort the indicators according to level of correlation

Take the first eight indicators of a relatively large correlation.

Predict future value of indicators according to Grey Model. The pre-expo stage of the 2010 expo is 2004~2009. Predict 2004 to 2009 data using gray GM (1,1) model and 1999~2003 data. Then we get the Expo incremental by minus the predicted value from the actual value. Then we can calculate the growth of the incremental.

## 6 The Expo Stage

Do regression using data form May- October 2010, calculated the third medium-term impact of industrial added value of individual indicators

## 7 The After-Stage of the Expo

For discussion of the economy after the World Expo can be further divided into two stages, namely the long-term and short-term. Short-term effect is similar to the Olympic Games' trough effect. ""Trough effect" refers to the Olympic Games-related end of the investment when the event is in its coda, so the opening of the Olympic Games special needs has largely disappeared, the economy is going into recession. Expo will also present trough effect, according to previous experience of the Expo, there may be lack of demand, the tourism industry issues such as depression, it is due to the excessive release of purchasing power in a short time.

Long-term effects after the Expo stage during the World Expo is a means of increasing investment in infrastructure, optimize the industrial structure and improve the host city of World Expo popularity and international reputation of the factors in the long term will have a positive impact on the region's economy. And this long-term impact, scope still difficult to measure directly the characteristics of the Expo is still the case more difficult to quantify, can not set a specific time frame space.

## 8 Quantitative Assessment of Stages of the Economic Cycle of the World Expo Impact

### 1) Pre-expo stage

Multi-index comprehensive quantitative analysis by Entropy method

Get the right weight by Entropy Method and calculate the comprehensive evaluation index  $V_i$  of year i.

We see the Combined economical effects of expo through the years. The Pre-expo stage influential score is  $V_6 = 72.65\%$

In the results we can also find that the influence on regional economics of the Expo shows a Ladder-like growth: fast growth from 2004 to 2005, the growth slowed to a halt, then the growing speed reached its original level., and eventually reaches its maximum in 2009, this value is the whole influence of the former Expo stage of economic development function

2) The expo stage

The influence on the tertiary industry:

Calculated the influence of composite indicators of each month in Expo period

The result indicates that the influence of expo is decreasing (though still stays at a high level) as the event comes close.

3) The after-stage of expo

According to the influence of changes in the value trend, impact indicators rate is not constant, but a linear decreasing function of time. Judged by the common sense, the influence of the value until the end with the World Expo will be conducted, and finally the influence of close to 0. Influence is not unlimited increase in value, but ultimately tends to a constant.

## 9 Promotion of the Model: Beijing 2008 Olympic Games

2008 Beijing Olympic Games and Shanghai World Expo have a lot in common. First, they are both worldwide event. Besides, Beijing and Shanghai are China's two largest cities with relative concentration of resources, and play an important role in regional economics.

Therefore, we can adapt the quantitative assessment methods described above to assess on the influence on the regional economy of Beijing Olympic Games, and compared with the Shanghai World Expo. This could reflect the domestic impact of major economic events.

Compare the regional economical influence between 2008 Beijing Olympics and 2010 Shanghai Expo.

The results show that in the aspect of regional economical influence, the two events had little difference. The 2008 Beijing Olympics has a slightly larger influence on regional economics than the 2010 Shanghai Expo.

## References

1. Shanghai Municipal Statistics Bureau, Shanghai Statistical Yearbook (September 12, 2010), <http://www.stats-sh.gov.cn/2004shtj/tjnj/tjnj2010.html>
2. Shanghai Tourism Bureau, the Municipal Tourism Bureau information catalog (September 12, 2010), [http://lyw.sh.gov.cn/message/information/content/gov\\_bulletin\\_new3.html](http://lyw.sh.gov.cn/message/information/content/gov_bulletin_new3.html)
3. Zhu, D., Zhou, X.: Shanghai World Expo will impact the quality of life and countermeasures. In: Shanghai World Expo. Impact and Response, pp. 3–49. Tongji University Press, Shanghai (2004)
4. Gregory, N.: Mankiw, Macroeconomics, pp. P249–P252. China Renmin University Press, Beijing (2005)

5. Shanghai statistics, monthly data, 9 (February 12, 2010), <http://www.stats-sh.gov.cn/2005shtj/sjfb/ydsj.htm>
6. National Bureau of Statistics, China Statistical Yearbook, XXXXX (2009), <http://www.stats.gov.cn/tjsj/ndsj/2009/indexch.htm>
7. Wang, K., Song, H.: Three objective weighting method weights comparative analysis of technical and economic and management research P48-P49 (March 2003)

# An Optimal Model for Repeater Coordination

Shiyun Yu<sup>1</sup> and Hao Li<sup>2</sup>

<sup>1</sup> Jingsiyuan 15-607 Shanghai University of Finance & Economics, 777 GuoDing Road, YangPu District, Shanghai Province, China

<sup>2</sup> North 4 dormitory building, 238, Shanghai University of Finance & Economics, 777 GuoDing Road, YangPu District, Shanghai Province, China

**Abstract.** Because of the limitation from line-of-sight transmission and reception, there is a necessity of repeaters to pick up weak signal, amplify them and retransmit them on a different frequency. However, two nearby repeaters interfere between each other when they transmit the signals in similar frequencies.

Our models find the minimum number of repeaters under certain limitations as the introduction mentioned in the paper.

We have made substantive conclusions by methods used as follows.

The Monte Carlo simulation has been perfectly used in our models, and the simulation results are consistent with the result worked out from our model.

Uses several techniques while modeling such as optimization and substitute thoughts in a complicated system.

We utilize simulation technique to test our model. It demonstrates that our method is robust.

**Keywords:** optimization, computer simulation, Monte Carlo, Repeater coordination.

## 1 Introduction

To solve this problem there are three methods available:

- The distance between two repeaters is long enough that the coverage range of the two repeaters don't have intersection.
- Two nearby repeaters (they have intersection) transmit on sufficiently separated frequencies. (when we assign frequency point to two nearby repeaters they don't share the same.)
- Two nearby can share the same frequency pair with the CTCSS technology.

Continuous tone-coded squelch system (CTCSS) associates to each repeater a separate subaudible tone, thus a repeater can only respond to received signals with its specific PL tone.

- In this research, our goal is to make model to determine the minimum number of repeaters to accommodate 1000 simultaneous users in a flat area of 40 miles radius. And optimize the model.

The available spectrum 145MHz to 148MHz is belong to amateur radio frequency. [1] Thus our research is on the Amateur radio aspect.

In addition, we will discuss the two additional scenarios:

- There are 10,000 users

## 2 Nomenclature

Symbol	Meaning
$D_{cd}$	LOS distance between two direct users
$R_{LOS}$	Coverage radius of a repeater
$R_e$	The radius of earth
$N$	total population
$R$	radius of total area
$T$	total number of repeaters needed
$S'$	area of the regular hexagon
$k$	number of repeaters needed in one hexagon (small circle)
$r'$	side length of the hexagon
$r$	radius of circum circle of the regular hexagon
$\rho$	density of population

## 3 Basic Assumptions

- We only consider the line-of-sight propagation of VHF signal. Scattering, diffraction and inflection of VHF radio are ignored. Besides, we consider earth to be a perfect sphere with a radius of  $R=6.38 \times 10^3$  km.
- A repeater picks up VHF radio spectrum on a certain frequency, while a user sends a spectrum at certain frequency during one transmission. If the two frequencies meet, the repeater could transmit the signals sent by user. In other words, a repeater allows only one user to transmit VHF radio on certain frequency at a time.
- There is no obstacle on the ground. Nothing can stop the wave unless it reaches the ground or be received by the other users or repeaters.
- In the flat area the highest thing is the repeater.
- Within the area, the signal transmitted by the initial transmitter is the only radio wave source and its frequency is between 145MHz to 148MHz. It means there is no additional radio wave from other things in the space.
- Two people can connect with each other directly without using repeaters if they locate within a distance small enough.
- All the two-way radios are low power users. The distance of direct user-to-user contact has a limitation.

The distribution of the users in the circular flat area is a uniform distribution.

## 4 The Calculation of Line-of-Sight Distance

In wireless channels, not only does radiation loss occur, but also one antenna may not "see" another because of the earth's curvature.[2] Since the VHF spectrum travels straightly via air and we can regard the earth as perfect sphere, there is a range of LOS distance depends on the height of two antennas.

LOS is the short of line-of-sight, *LOS distance* is the length of tangent line between two antennae.

## 5 Two Models

### Broadcast Model

#### Establish Coverage Model

In this situation, we should build the minimum number of repeaters in areas, so that all the repeaters can cover the whole area. Any one of the 1,000 users can transmits a signal and the signal can be received by all the other users via repeaters.

In mathematics aspect, our goal is to use the minimum number of small circles (regular hexagons) to cover the big circle (with a radius of 40miles).

#### Solution on Optimization of the Model

##### ◆ optimize the location of the repeaters

In the coverage model we put a repeater in the center of each small circle (regular hexagon)[3], so there are 31 repeaters in the big circle. However, to cover the whole area, another 6 repeaters is needed. To optimize this defect, we adjust the outside points into the big circle. The minimum number of repeaters needed for a broadcast model is 37 in the big circle.

##### ◆ allocation of frequency points

Since the nearby repeaters have a part of overlap, if the two repeaters share the same frequency without PL, they will interfere with each other. Thus, we need allocate frequency points to each repeater, and if there is not enough frequency points we can use PL tone.

## 6 Tow-Way Radio Model

### Establish Model

A two-way radio is a radio that can both transmit and receive (a transceiver), unlike a broadcast receiver which only receives content.[4]

Because of the characteristic of the two-way radio(P2P), if all the users want to communicate simultaneously, the number of repeaters will be very large( for 1000 users approximately 500 repeaters). Therefore, it is very hard to achieve direct repeater-to-repeater communication without huge interference. Additionally, in practice, HAM

(radio amateur) would not do repeater-to-repeater direct communication. According to relative documents, when connecting one repeater to others, hams commonly use systems providing special techniques such as IRLP, ILINK, ECHOLINK and so on.[5] Thus in this two-way radio model we assume that repeaters cannot communicate with each other, which implies that the communication range for a user is limited: one user can only reach users within LOS distance of available repeaters.

To avoid mutual interference, frequency point intervals for different repeaters should no longer be 0.6MHz. They should be reallocated to the repeaters so that spectrum transmitted by one repeater would not be received by another one.

Take 700KHZ to be the interval, we get 5 frequency points:145MHZ, 145.7MHZ, 146.4MHZ, 147.1MHZ, 147.8MHZ.

In a small circle we will put several repeaters in the center to meet the need of the users within the range of the repeaters. Since these several repeaters have the same coverage radius ( $R_{LOS} = 8(miles)$ ), we regard these several repeaters as a combination when calculate their control rage.

### Non-linear Programming of the Model

There are 5 frequency points “A,B,C,D,E” (see figure) and 54 PLs, to avoid interference, nearby repeaters could not share one frequency points with the same PL. While sharing one frequency points of different PL or share one PL within different frequencies is allowed. Thus there are  $5 \times 54 = 270$  strategies from which each repeater could choose from. We name the 270 strategies ‘FL’.

● *Three restrains:*

- a) Effective coverage of the repeaters should cover the whole area.
- b) Requirements of simultaneous user number should be met. In the worst case, every user needs a repeater to connect to each other.
- c) There are 270 ‘FL’s. Adjacent repeaters should not share one FL.

The first restrain is similar with the first model, which can be viewed as a coverage problem. However, in this model, because of the relativity between population and FL assignment, we need to control  $r$  to meet the limitation of the number of FL.

Take  $r$  as variable, we could establish a Non-Linear Programming problem.

One assumption is that the distribution of the users in the circular flat area is a uniform distribution. Thus we can introduce the parameter: density of population, to indicates the number of people for each square miles.

$$\rho = \frac{N}{\pi R_0^2} \tag{6.2.1}$$

$$r' = r \tag{6.2.2}$$

$$n = \rho \cdot S' = \rho \cdot \frac{3\sqrt{3}}{2} (r')^2 \tag{6.2.3}$$

$$\delta = \frac{\pi D_{cd}^2}{S_{LOS}} = \frac{D_{cd}^2}{R_{LOS}^2} \tag{6.2.4}$$

The number of repeaters needed in one regular hexagon is

$$k = \left\lceil \frac{n}{2} \right\rceil = \left\lceil \frac{3\sqrt{3}}{4} \delta \rho (r')^2 \right\rceil \tag{6.2.5}$$

Total number needed in the area is

$$T = \rho k = \frac{\pi R_0^2}{S'} \left\lceil \frac{3\sqrt{3}}{4} \delta \rho (r')^2 \right\rceil \tag{6.2.6}$$

***Our Goal is to Minimize T***

One more restriction for programming which cannot be ignored is the allocation of frequencies and PLs:

- ◆ One regular hexagon can at most contain 90 FLs

According to the deduction of four color theorem, we can just use three kinds of colors to color this cellular net, so that two nearby hexagons have different color.[6]

Applied for this mathematics theorem, since two nearby small circle cannot have the same FL, as it shows, we can group all the FLs into 3 groups (group A,B,C or group 1,2,3)

Thus the number needed in each one of the regular hexagon cannot exceed the number 90 (  $270 \div 3 = 90$  ).

- ◇ Other restrains:  $r$  could not exceed the LOS distance. Besides, repeaters shouldn't be too near to each other, or there may be a interference  $3 < r \leq 8$ .

***Solution of the Model:***

Solving the Non-Linear Programming problem, we can get the results:

$$r = 7.93956, T = 430, k = 14 < 90$$

Thus, the total number of repeaters should be 430.

**10,000- User Condition**

**Broadcast Model on 10,000-User**

For broadcast model, as long as the entire area is covered, every users could receive signal and any one can send signal to all the others. Thus there is no difference between the situations with 10000 users and with 1000 users. Model 5.1 and model 6.1 have the same result: the minimum number of repeaters is 37.



### Two-Way Radio Model on 10,000-User

However, it is not true with the two-way radio model. As the number of user increases, the density of people  $\rho$  increases, making a great difference.

Resolve the programming problem, we get the solution:

$$r = 6.3658116, T = 4297, k = 90$$

In this case, the numbers of frequencies and PLs is a tight restrain.

Now we reconsider the inscribed square strategy. Because compared to inscribed regular hexagon strategy, it has a even looser restrain on the maximum number of FLs one station can hold.

We can just use two kinds of color to color it, so that the nearby squares are in different color. Thus two sets of different FLs can satisfy the FL constrain.

$$k \leq \frac{270}{2} = 135$$

$$r' = \sqrt{2}r$$

$$S' = 2r^2$$

Applied for the programming problem, the solution is  $r = 8, T = 4320, k = 110$ ;

As we can see,  $4297 < 4320$ , so the inscribed regular hexagon strategy still outperform the inscribed square strategy.

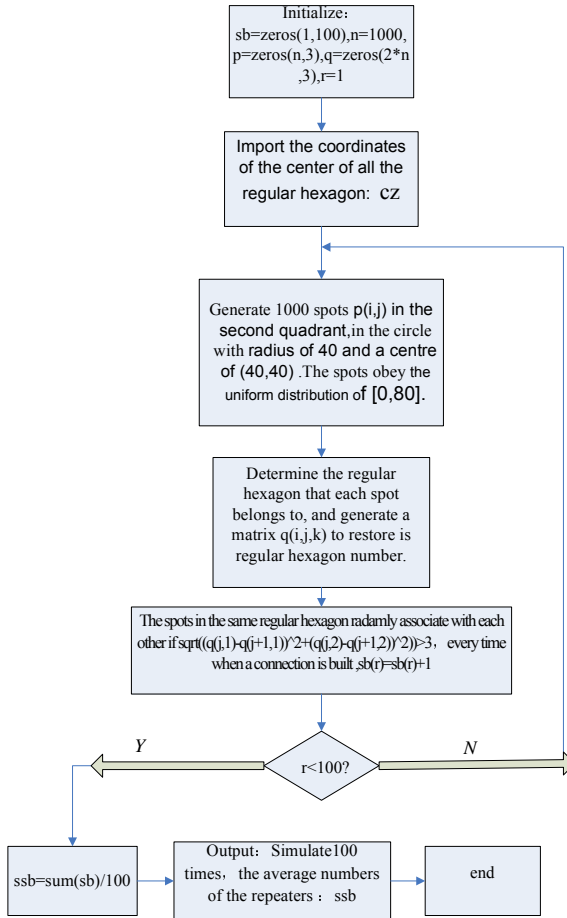
## 7 Model Test: A Simulation Model

### Model

The two-way radio model we established before is based on population density  $\rho$ , which is an index of average value. As supplementary, a simulation model is developed.

Simulating rules:

- ① The coordinate of the 1,000 simultaneous users are in a circle with the radius of 40 miles and Obey the uniform distribution of  $[0,80]$ .
- ② Every repeater in the centre of the cell can cover a space with a certain radius (7.93956 miles or 6.3658 miles).
- ③ If the distance between two users is less than 3 miles, they do not need a repeater to connect.;



**Fig. 1.** Flow chart of the simulation

Explanation of symbols:

- (1) cz: the coordinates of the center of the regular hexagon.
- (2) n: number of users.
- (2) p(i,j): the Coordinates matrix of the 1000 random users.
- (3) q(i,j,k): the Coordinates matrix of the 1000 random users and the regular hexagon they belongs to.
- (4) sb: the numbers of the repeaters simulated 100 times.
- (5) ssb: average number of the repeaters.
- (6) r: how many time we simulate.

## 8 Sensitivity Analysis and Conclusion

- In broadcast model: the height is vital: if the height of repeater increase by 1, LOS will change 0.473, and number would change by.
- In two-way radio model, the sensitivity is given in shadow price (Lagrange multiplier).

When user number is 1000, Lagrange multiplier is 0, which means if the limits for 'FL' increase by 1, T would decrease by 0. While in the 10000-user case, the repeater needed for Lagrange multiplier is 47.73444, which means the frequency and PL number is the main limit for better solution. If the limits for 'FL' increase by 1, total number needed would decrease by 47.73444.

LOS distance does not weigh so much in the two-way radio model as it does in the broadcast model, for its price of Lagrange multiplier is both 0, which means LOS distance is not a important limitation in the two-way radio model.

These results perfectly fit our intuition. In broadcast model, only one person use the repeater, the main constrain is the coverage; while in two-way model, there are so many users using the repeater that the coverage becomes less of a problem and the main constrain is how many people the system can carry.

## 9 Future Work

Since the weakness we mentioned above, if we want more accurate data and parameters to work out a more accurate result, we should do more research about the condition. (In this model, we assume that it is ideal condition.) A more stable programming should be achieve by other techniques, since the non-linear programming is not stable.

## References

1. Wikipedia, the free encyclopedia (2011), [http://en.wikipedia.org/wiki/Amateur\\_radio\\_frequency\\_allocations](http://en.wikipedia.org/wiki/Amateur_radio_frequency_allocations)
2. Johnson, D.: Under a Creative Commons Attribution License (CC-BY 1.0). Open Educational Resource. Last edited by Connexions on June 9, 5:43 pm GMT-5 (2009), <http://cnx.org/content/m0538/latest/>
3. Wikipedia, the free encyclopedia (2009), [http://en.wikipedia.org/wiki/Branch\\_and\\_bound](http://en.wikipedia.org/wiki/Branch_and_bound)
4. Wikipedia, the free encyclopedia (2010), [http://en.wikipedia.org/wiki/Two-way\\_radio](http://en.wikipedia.org/wiki/Two-way_radio)
5. Website (February 6, 2011), <http://www.ac6v.com/repeaters.htm>
6. Wikipedia, the free encyclopedia (2011), [http://en.wikipedia.org/wiki/Four\\_color\\_theorem](http://en.wikipedia.org/wiki/Four_color_theorem)
7. Jiang, X., Yang, M.: Communication Theory, pp. P250–P255. The university of Haerbin (March 1, 2010) ISBN:9787560329925

# Copyright Self-registration and Its Secure Authentication Based on Digital Watermark

Jianbo Liu and Cheng Yang

Information Engineering School, Communication University of China,  
Beijing, China

**Abstract.** The copyright registration and authentication is an important task in copyright protection system for the digital work. The traditional centralized copyright registration and authentication system is less security and not fit for large quantities of digital work. In this paper, a novel copyright self-registration mode and authentication scheme is proposed for the copyright protection. In this scheme, a secure copyright protection authority and the copyright watermark of the owner are used. After described the architecture of the copyright protection authority, the secure copyright self-registration protocol based on the public elliptic curve and session key is proposed. Based on the proposed architecture and the copyright watermark, the owner identification protocol and the ownership verification protocol are all discussed in detail. After tested in the IPTV system and analyzed in the aspect of security, the proposed scheme and protocols are high performance and secure.

**Keywords:** copyright registration, authentication protocol, copyright protection authority, replacing attack.

## 1 Introduction

Protecting the copyright of the digital work is a common sense which already demonstrated in the World Intellectual Property Organization Copyright Treaty, in the Copyright Law [1], and in the Regulations on the Protection of Computer Software [2], etc.

At present, for the copyright protection of the digital works, the centralized copyright registration mode is widely used [3][4]. But because the bottleneck of the registration server and the complexity of the register process, this mode is appropriate to authors who owns few works. These authors submit final products and abstracts of their digital works to copyright protection authority (CPA). The CPA checks the validity of the submission and signs the ownership of the copyright. Furthermore, via the public portal of exhibition and business in the CPA, the registered and protected digital works can be distributed to the public. In this copyright registration mode, the CPA is the agent or spokesman of the authors.

For the organization, the group or the enterprise who own lots of digital work, the copyright self-registration mode based on the watermark is proposed in this paper.

The digital copyright information based on the digital watermark [5][6] can be used to implement the copyright registration and authentication. The author's copyright watermark can be treated as the owner's fingerprint. Before the abstract or the product is shown or distributed, the owner's fingerprint can be embedded into them using the watermarking technique. When some dispute occurs, the owner or the law enforcement agencies can submit the seized products to the CPA. The CPA will identify the owner or verify the ownership.

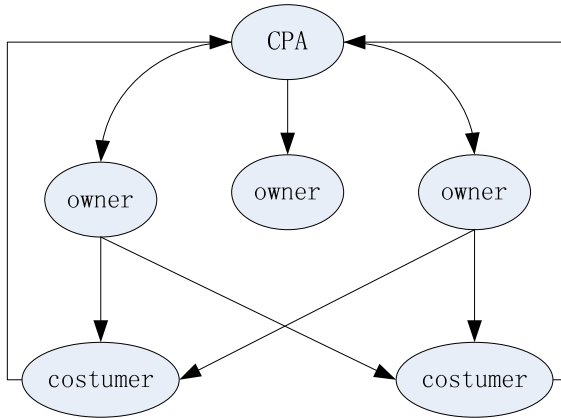
In this paper, a novel copyright registration and authentication based on the digital watermark is proposed. In section 2, the copyright self-registration mode and the architecture of the copyright protection authority are designed. In section 3, the copyright self-registration protocol is proposed. In section 4, the owner identification protocol and the ownership verification protocol are discussed. At last section, the security of the proposed scheme and protocols are analyzed and then implemented into the IPTV system.

## **2 Copyright Self-registration Mode and Architecture**

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

### **2.1 Copyright Self-registration Mode**

For the organization, the group or the enterprise who owns lots of digital works, the copyright self-registration mode based on the digital watermark is proposed in this paper. These owners will directly face to the customers and will have their own copyright protection platforms in which the protected digital works and customers' information will be stored and managed [7]. These owners will have their own portals of exhibition and business. In the copyright registration mode, the CPA acts as the trust third-party and is the authentication center (AC) which is shown in the figure 1. Through the interaction between one of the owners and the CPA, the copyright information is checked and signed by the CPA, and then the digital works and copyright information are encapsulated by the owner. The middle products and related information are stored and managed by the owner himself. When the middle products are sold to a customer through the business portal, the CPA will participate in the process to sign customer's fingerprint and ensure the security of the transaction and the privacy of the customer, the customer's fingerprint will be embedded into the middle works by the owner in order to support the piracy tracking, at the same time, the owner cannot counterfeit the fingerprint.



**Fig. 1.** The relationship in copyright self-registration mode

In the copyright self-registration mode, the CPA and the owner are all have valid digital certificates and are respectively in charge of:

- 1) The owner collects and manages its original digital works and middle products.
- 2) The owner generates the copyright watermark which will be signed by the CPA.
- 3) The owner embeds the signed copyright watermark into digital works which is encapsulated to be middle product or watermarked product.
- 4) The owner takes charge of the management of middle products. If the product is lost, the piracy root is the owner itself.
- 5) The owner could establish the exhibition and business platform.
- 6) During the transaction between the owner and the customer, the customer's fingerprint is signed by the CPA and embedded into middle product by the owner himself. The final product is then distributed to the customer.
- 7) The CPA provides the copyright authentication platform to identify the owner of digital works, to tell the true or the false of the ownership.
- 8) The CPA provides the piracy identification platform to identify the owner's piracy and the customer's piracy.

The copyright self-registration mode demands the owner has capability of data processing with high performance. In order to embed copyright and fingerprint watermark, the robust digital watermark algorithm packages are also needed.

## 2.2 Architecture of Copyright Protection Authority

The CPA is the kernel of the proposed scheme. It is a fair and trust third party. It acts as the root of trustiness chain. The CPA is divided into four parts which are the registration, exhibition, transaction and verification. In this architecture, the related entities are the CPA, the author, the customer, the valid-requester and the pirate. In the figure 2, the architecture of the CPA is shown.

The author is the producer of the digital work or the copyright owner. He actively submits application of copyright and digital work information to the CPA. Then the author embeds the copyright watermark into the digital work.

The valid-requester is an entity who doubts the copyright of some digital works and submits verification application.

The pirate can be independent entity or be badness entity with legal identity, for example the pirate is same as the customer.

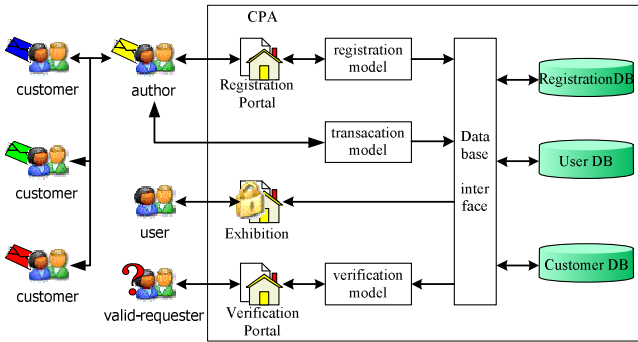


Fig. 2. Architecture of copyright protection authority

In figure 2, the registration module is in charge of the copyright registration. The generation and the signature of the copyright watermark of the digital work, the storing of the abstract or product information are all its tasks.

The transaction module is in charge of the generation of customer fingerprint according to information from the author to the customer. The customer information will be stored in customer database.

The verification module is in charge of the copyright verification of the digital work. The right verification of valid-requester, the extracting of copyright watermark or fingerprint of abstract or product, the verification of the copyright, the tracking of the pirate root are all its tasks.

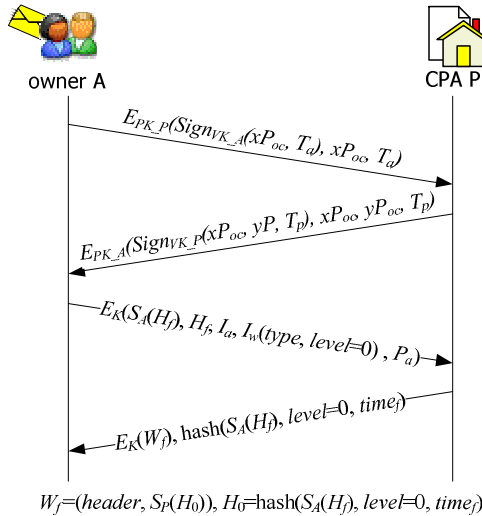
### 3 Copyright Self-registration Protocol

The copyright registration is the base and the first step of the digital works protection system. Only the registered digital works can be protected and only its owner can experience the protecting service. The copyright self-registration protocol is shown in figure 3.

In figure 3, the  $S(.)$  is a signature function, such as the RSA. The  $E(.)$  is an encryption function, such as the AES. The prerequisite of the copyright self-registration is that the register-requester has the ownership of some digital works, the digital works should not have been registered before, and the owner A and the registration module P in CPA all have the public keys from CA. During the process of copyright self-registration, the key is how to secure the interaction between the owner

and the CPA. The copyright information must be signed by the CPA and be encapsulated by the owner. Before the copyright self-registration protocol, the owner A must get the watermark embedding component offline from the CPA.

Before the protocol, the owner A and the CPA P will get their public keys and private keys [8] from general certificate authority (CA) system. These keys are  $(PK_a, VK_a)$  and  $(PK_p, VK_p)$ . At the same time, a  $P_{oc}$ , on public elliptic curve, is shared between the owner and the CPA.



**Fig. 3.** Copyright self-registration protocol

1. The P and the A authenticate each other, and they negotiate the session key K:

1) The owner A generates a secret random number  $x$ , computes the  $xP_{oc}$ , signs the  $xP_{oc}$  and the processing time  $T_a$  using  $VK_a$ . The signed information and  $(xP_{oc}, T_a)$  are encrypted together with  $PK_p$ , and then are sent to the CPA.

2) The CPA receives the message from the owner A and decrypts it with the  $VK_p$  to get the  $(xP_{oc}, T_a)$ . With the  $(xP_{oc}, T_a)$ , the CPA verifies the signature using  $PK_a$ . If the signature is valid, the CPA generates one secret random numbers  $y$ , computes the  $xyP_{oc}$ , the result is the session key  $K$ .

3) The CPA computes the  $yP_{oc}$ , signs the  $(xP_{oc}, yP_{oc})$  and the processing time  $T_p$  using  $VK_p$ . The signed information and the  $(xP_{oc}, yP_{oc}, T_p)$  are encrypted together with  $PK_a$ , and then are sent to the owner A.

4) The owner A decrypts the message into  $(xP_{oc}, yP_{oc}, T_p)$  using  $VK_a$ , verify the signed  $(xP_{oc}, yP_{oc}, T_p)$  using the  $PK_p$ . If the signature is valid, the owner A computes the  $xyP_{oc}$ , the result is the session key  $K$ .

5) The owner A encrypts the  $(yP_{oc}, T_p)$  with the  $K$ . The result is sent to the CPA. If the  $K$  is generated successfully, the CPA can decrypt the message and confirm the key generation.

2. The A computes the hash value  $H_f$  of digital work  $P_f$ , signs the  $H_f$  to be  $S_A(H_f)$  and then encrypts  $(S_A(H_f), H_f, I_a, I_w(\text{type}, \text{level}=0), P_a)$  using  $K$  to be  $E_K(S_A(H_f), H_f, I_a,$



$I_w(\text{type}, \text{level}=0)$ ,  $P_a$ ). The result is sent to the P.  $I_a$  is the author's information,  $I_w(\text{type}, \text{level})$  is the digital work information. The  $\text{type}$  is audio, video, film, software or other types in the copyright laws. The  $\text{level}$  is the integrality level of the digital work.  $\text{level}=0$  means that the registered content is the digital work itself.  $\text{level}=1$  means that the registered content is the abstract of the digital work.

3. After received the message, P use session key  $K$  to decrypt the message, and use the public key  $PK_a$  and  $H_f$  to verify the signature  $S_A(H_f)$ , if the authentication is passed, then you can enter the next step, otherwise, quit the copyright registration process, and return to A the "Authentication failed" message.

4. Through a database interface, P finds  $H_f$  in the Registration DB to confirm whether the copyright of work has already registered, if no  $H_f$  in the database, you can enter the next step, otherwise, quit the copyright registration process, and send message "re-registration" to A.

5. If the  $P_a$  is not empty, then:

1) P generates copyright watermark  $W_a$  from the abstract of the digital work:

$$W_a = (\text{header}, S_P(H_1)), H_1 = \text{hash}(H_a, \text{level}=1, \text{time}_a), H_a = \text{hash}(P_a),$$

in which the  $\text{header}$  is the pseudorandom sequence copyright identifier which is used to imply the watermark is already embedded, the  $\text{time}_a$  is the timestamp for the abstract from any security time source,  $S_P(\cdot)$  is the signature function for P.

2) P calls the copyright watermark algorithm package to embed the copyright watermark  $W_a$  into the  $P_a$ , and receives the  $P_{aw}$ .

6. P generates the copyright watermark  $W_f$  of the digital work:

$$W_f = (\text{header}, S_P(H_0)), H_0 = \text{hash}(S_A(H_f), \text{level}=0, \text{time}_f),$$

and encrypt the  $W_f$  using the session key  $K$  to be the  $E_K(W_f)$ .

7. Through a database interface, P stores  $H_0$ ,  $H_1$ ,  $H_a$ ,  $H_f$ ,  $S_A(H_f)$ ,  $I_a$ ,  $I_w$ ,  $\text{time}_a$ ,  $\text{time}_f$  and  $P_{aw}$  into the registration database, and sends  $(E_K(W_f), \text{hash}(S_A(H_f), \text{level}=0, \text{time}_f))$  to A;

8. After received the message, A decrypts the message with the session key  $K$ , gets the  $W_f$  and  $H_f$ , and verify  $H_f$  using the public key  $PK_P$  and the  $\text{hash}(S_A(H_f), \text{level}=0, \text{time}_f)$ , if the authentication is passed, then success to confirm the copyright information, or else sent the message "Authentication failed" to P.

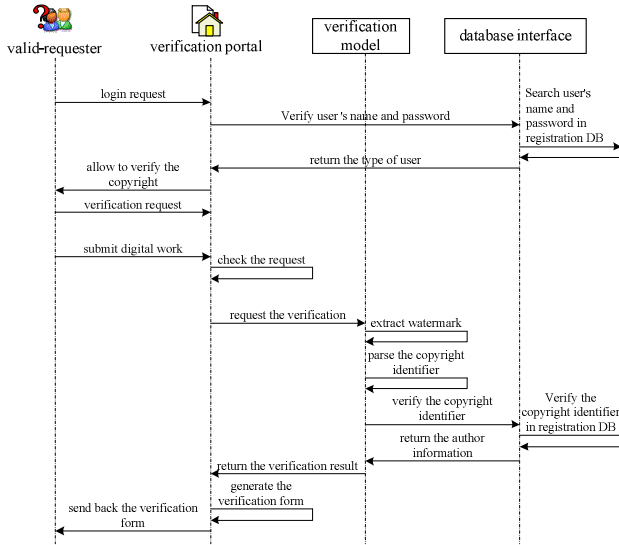
9. A embeds the copyright watermark  $W_f$  into the digital works  $P_f$  using the copyright watermark embedding machine  $f(\cdot)$ , receive the middle product  $P_{fw}$ , and complete the copyright registration.

## 4 Copyright Authentication of Digital Work

The copyright verification of digital work is fulfilled by the Copyright Protection authority (CPA). The copyright verification includes the identification of the owners of digital works and the verification of the ownership.

### 1. The owner identification protocol

The request of the owner identification can be put forward after the digital work registration, or can be come up with by custom or author when they need to identify the owner of copyright. The following figures show the owner identification protocol process.



**Fig. 4.** Identify the owner of the digital work

- 1) The valid-requester gives his user name and password, and submits verification application.
- 2) Through database interface, the verification portal can identify validity and the type of user, make sure he is a valid user.
- 3) The valid-requester is allowed to submit the request for verify the copyright of digital work, and submit the digital work  $P_{fw}$  which need to be verified.
- 4) The verification portal reviews the request, including the correctness of application, the consistency between application and digital work.
- 5) If review failed, the verification portal returns an error message to the valid-requester. Otherwise, the valid requester can submit the digital work to complete the verification process.
- 6) The verification module uses the digital watermark algorithm to extract the copyright information  $W_f$  from the  $P_{fw}$ , and then decode the  $W_f$  to get the *header* and the  $H_0$ .
- 7) If the *header* is not correct, the digital work is considered without the copyright registration.
- 8) If the header is correct, through the database interface, the verification model is to verify the  $H_0$  in the registration DB, and to obtain the corresponding digital work information including  $S_A(H_f)$ ,  $H_f$ ,  $I_a$ ,  $I_w$ ,  $time_f$ . If the verification failed, the digital work information is *null* which indicates the digital work without the copyright registration.
- 9) The verification model returns the verification result to the verification portal. The results include whether the digital works is registered or not, author information of digital works, etc.
- 10) The verification portal generates the result form and sends back the form to the valid-requester.

## 2. The ownership authentication protocol

If the copyright dispute arised, the CPA is to authenticate the ownership of the digital work. Any party in copyright dispute may submit a request for verification of ownership. The figure 5 shows a schematic of the ownership authentication of the digital work.

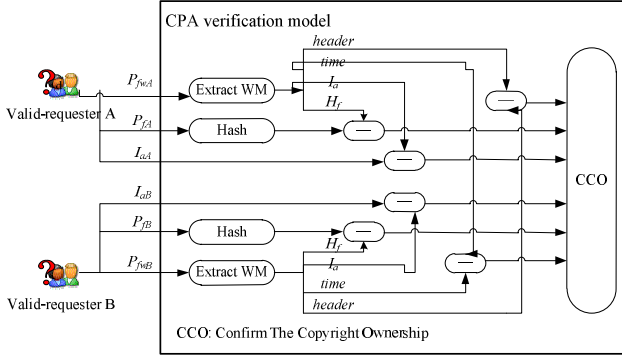


Fig. 5. Verify the ownership of the digital work

The ownership authentication is fulfilled by the verification model in the CPA. After the copyright information  $W_f$  is extracted from the digital work, and then under the condition of a valid *header*, the ownership is verified by determine the consistency of the original digital work's hash value and the pre-stored hash value, by determine the consistency of the author's information and pre-stored author's information. If all parties has the ownership of the digital work, the time-stamp will be used for determine the ownership. The process is shown as follows.

1) The two parties A and B in the copyright dispute respectively authenticate with the verification model in the CPA and obtain the session key  $K_A$  and  $K_B$  using the same protocol in the section 3.

2) A and B respectively submit their information  $I_{aA}$  and  $I_{aB}$ , the similar protected digital work  $P_{fwA}$  and  $P_{fwB}$ , the original digital work  $P_{fA}$  and  $P_{fB}$  to P. A and B respectively send the message  $(E_{K_A}(P_{fwA}, P_{fA}, I_{aA}))$  and  $(E_{K_B}(P_{fwB}, P_{fB}, I_{aB}))$  to P.

3) After the message is decrypted using the session key  $K_A$  and  $K_B$ , P extracts the copyright information  $W_{fA}=(header_A, H_A)$ , and  $W_{fB}=(header_B, H_B)$  from  $P_{fwA}$  and  $P_{fwB}$  using the copyright watermark extraction algorithm.

4) Set the initial ownership set is  $S=\{A, B\}$ .

5) According to the *header*, the ownership of the digital work is determined:

If the  $header_A$  is valid,  $header_B$  is invalid, then  $S=S-\{B\}$ .

If the  $header_A$  is invalid,  $header_B$  is valid, then  $S=S-\{A\}$ .

6) If  $S= \emptyset$ , then the digital work is without the copyright registration, the ownership is not belong to A or B. The process of the ownership verification is terminated.

7) If  $S \neq \emptyset$ , for each element of  $S$  (denoted as  $D$ ), we determine whether it has the ownership of the digital work:

The  $H_D$  is decomposed from the  $W_{fD}$ . Through the database interface, we search the  $H_D$  in the registration database. If no record, then  $S=S-\{D\}$ , otherwise, we access  $H_f$ ,  $I_a$  and  $time_f$  from the registration database, and continue.

The hash value of the original digital work  $P_{fD}$  is calculated, if the hash value isn't equal to the  $H_f$ , then  $S=S-\{D\}$ , or else continue.

If the author information  $I_{aD}$  and  $I_a$  are inconsistent, then  $S=S-\{D\}$ .

8) If  $S=\emptyset$ , then the digital work is without the copyright registration, the ownership is not belong to A or B. The process of the ownership verification is terminated.

9) If  $S=\{A, B\}$ , according to the time stamp which is obtained from the registration database, if  $time_{fA} < time_{fB}$ , then  $S=S-\{B\}$ , otherwise  $S=S-\{A\}$ .

10) P returns the results of the ownership verification to each element in S.

## 5 Security Analysis and Test

### 1. Implementation and Test

The proposed scheme and protocols are implemented for IPTV VOD service of video product.

In this environment, the VLC streaming server and RTSP protocol are adopted in owner. Video contents distributed through the IP network to the STB (Set Top Box). The CPA is a standalone server with the access control using the firewall. It is in charge of the registration of video product, the generation of copyright watermark, and the copyright authentication. In STB, the SoC Broadcom 7405 and the embedded Linux are adopted. The test stream is a 8Mbps mpeg-2 video.

In the owner, the copyright watermark embedding module is a part of the pre-packaging server. In the STB terminal, the copyright watermark extracting module is running. The watermark modules use the fast DFT based watermarking algorithms. These types of algorithms are high performance, robustness.

In table 1,2, the performance of the scheme is shown.

From the table 1-2, in the test system, very few time and CPU power are consumed and other processes, such as the internet browser, are not affected.

**Table 1.** Time Consumption of Watermark Algorithms

embed-time	extract-time
11Mbps	24Mbps

**Table 2.** CPU Consumption of Watermark Modules

	watermark module in pre-packaging server	watermark module in STB
CPU power consumption	15%	6%

## 2. Security Analysis

1) In the centralized copyright registration mode, copyright watermark generation and embedment is entirely completed by the CPA. The author cannot forge the copyright, cannot deny the registration. But this mode adds the load on the CPA, extend the time of registration. Moreover, once the CPA cannot work, it will result in failure to complete the registration. Because the CPA will direct access the author's original digital works, once the CPA failed, it will cause great economic losses. However, the advantages of the centralized registration are that the author has no technical requirements, the author who apply for registration only need to fill out the appropriate application form and submit their works. It is fit for authors and groups who have a small amount of copyright digital works such as artists, individuals and others.

2) Copyright self-registration is fulfilled by both the owner and the CPA. The CPA is only responsible for generating the copyright watermark. The specific watermark embedding is completed by the owner himself, which makes load balancing between the author and the CPA in registration to reduce the burden of the CPA. For each owner it shortens the register delay. As in this registration, CPA cannot access the original digital work, it enhance the security. Copyright self-registration requires the owner has hardware and software to support the watermark embedding, encryption and decryption, and has the public key certificates. It is suitable for the owners who have large quantities of digital works. The owners can be the manufacturers of contents, the copyright agents etc.

3) The copyright watermark of digital works have characteristics such as cannot be forged, non-repudiation, verifiable and uniqueness.

Because the owner cannot forge the signature  $S_p$  of the CPA, the owner cannot forge the copyright watermark (*header*,  $S_p(H_0)$ ) which is generated by the CPA. In addition, because the  $S_p(H_0)$  relates to the hash value of the original digital work, so if the owner embed the copyright watermark into other digital works, in the process of the copyright identification and the ownership verification, according to the difference between the hash value of the original digital work and the hash value which identified by the copyright watermark, it can further prevent the owner from forging the copyright. Because the  $H_0$  relates to the owner's signature  $S_A$ , you can search the registration database to determine the adscription of digital works' copyright, and the owner cannot deny that they own the ownership of the digital work. Non-forgeability and non-repudiation make the owners and their digital works to bind together.

In the ownership verification, by extracting and analyzing the watermark, we can obtain  $H_0$ , and carry out the ownership verification protocol. It can validate the adscription of the copyright by  $H_f$ ,  $I_a$  and  $time_f$  etc. When the two authors all have ownership of the same digital work, the ownership may be determined based on the time stamp.

Since the copyright information is generated based on the hash value and timestamp of a digital work, it is unique for any digital works.

4) The session key generated in copyright self-registration protocol is secure.

Service key secrecy: The proposed session key access protocol is based on the CDH (Computational Diffie-Hellman) problem, which is intractable on cryptographic elliptic curves. The session key  $K=xyP$  is not transmitted on network. Because the

only information about  $K$  could be intercepted is the  $xP$  and  $yP$ , the attacker could not derive the  $K$ .

**Session key freshness:** In the session key access protocol, the secret values  $x$  and  $y$  are randomly chosen by the owner and the CPA respectively, so each side can confirm the freshness of the agreed  $K$ .

**Session key confirmation:** the owner can confirm that the  $xP$  has been received by the CPA and can be used to compute the desired  $K$ . The CPA confirms that the  $K$  has been correctly derived by the owner.

**Resist the Replay attack:** In every protocol message, the time stamp is encrypted and signed. So if hacker intercepts messages and replays them in order to fake the CPA or the owner, the timestamp can be used to be an evidence of attack.

**Forward security:** At the end of the session key access, the secret number  $x$ ,  $y$  should be deleted. So after the validity period of a  $K$ , anyone who got all the parameters, even including the compromised private keys  $VK$ , cannot derive the session key. The date encrypted by the  $K$  in the past cannot be derived in future, which achieved a secrecy property of forward security.

## 6 Conclusion

In order to promote the informatization, the systematism and the internationalization process of the copyright trade, and for the requirement of the copyright protection, the digital works' copyright registration and verification are discussed. Because the traditional centralized copyright registration mode is not suitable for large quantities of digital works and is less security, the self-registration solution is proposed in this paper. In this solution, owners must have hardware and software support such as watermark embedding device, the public key certificate, etc. The burden of the system is kept balance, the registration delay is small. This mode is suitable for the owners who have a large amount of digital works. In order to support the copyright self-registration and verification, the CPA system architecture is designed which includes the registration portal and registration module, transaction module, exhibition platform, verification portal, verification module, security database access interface, and so on. Based on the established system architecture, the copyright self-registration protocol, the owner identification protocol and the ownership verification protocol are all designed and discussed. The copyright self-registration is fulfilled by both the owner and the CPA. The CPA cannot access the original digital work, it enhance the security. The copyright watermark have characteristics such as cannot be forged, non-repudiation, verifiable and uniqueness.

**Acknowledgments.** The work on this paper was supported by National Nature Science Foundation of China (60902061), the National Key Technology R&D Program (2008BAH28B06-05) and the 3rd phase of 211 Project of Communication University of China. The work is also supported by the Culture & Creative Industry Project of Beijing and the SARFT Scientific and Technological Project (2005-02-2).

## References

1. Copyright Law of China (October 2001), <http://www.bjcopyright.org.cn>
2. Regulations on the Protection of Computer Software (January 2002), <http://www.ccopyright.com.cn/crprotect>
3. Hu, Z., Gu, J.: Research on Media Information Copyright Protection Based on Security Container, pp. 26-31 (July 2002)
4. Yang, C.: Streaming media and problems about the protection of its digital right. *Net Security Technologies and Application*, pp. 82-85 (July 2003)
5. Xuebin, L., Zemin, L.: An image watermarking algorithm based on uniform multichannel subband transform. *Journal of Beijing University of Posts and Telecommunications* 26(2), 14–17 (2003)
6. Podilchuk, C.I., Delp, E.J.: Digital watermarking: algorithms and applications. *IEEE Signal Processing Magazine* 18(4), 33–46 (2001)
7. Qian, L., Nahrstedt, K.: Watermarking schemes and protocols for protecting rightful ownership and customer's rights. *Journal of Visual Communication Image Representation* 9(3), 194–210 (1998)
8. Katzenbeisser, S.: On the design of copyright protection protocols for multimedia distribution using symmetric and public-key watermarking. In: *Proceedings of the 5th International Query Processing and Multimedia Issues in Distributed Systems Workshop*, pp. 815–819. IEEE Computer Society Press (2001)

# A New Approach for SQL-Injection Detection

Cong-cong Shi, Tao Zhang, Yong Yu, and Weimin Lin

State Grid Electric Power Research Institute, Nanjing, 210003, China

**Abstract.** With the deepening of information construction, Web architecture is widely used in various business systems. While presenting convenience, these new technologies also introduce great security risks. Web security has been a serious issue of information security, and SQL-injection is one of the most common means of attack against Web services. SQL Injection often changes the structure of SQL statements. This paper proposed a self-learning approach to counter SQL Injection which can learn automatically the structure feature of all legal SQL statements to construct knowledge library based on SQL syntax tree in safe environments, and then match every SQL statement with knowledge library to find whether the structural feature has been changed in real environments. If successful, this SQL statement is legal. SQL statements which fail pattern marching are not determined as illegal immediately. Then, we take depth-feature check based on Value-at-Risk, and identity the true illegal SQL statements. This method which combines mode-matching and character-filtering can reach good results. Experimental results prove that this proposed approach holds good performance and perfect protection for SQL Injection.

**Keywords:** self-learning, syntax tree, pattern marching, feature filtering.

## 1 Introduction

With the development of Internet, dynamic business website development which is based on B/S model is more and more widespread. This model adopts platform architecture with ASP.NET+SQL+IIS, therefore more and more programmers use this model write applications. However, there are many unsafe factors in application system because of uneven programmers and immature security defense technology. SQL injection means a malicious attacker input deliberately structured data which contains malicious code. This executed code cause damage to the server or client. SQL injection is accessed by normal port and there is no difference between SQL injection and web page access, so the most common firewall can't detect it. SQL injection is the first of "Ten Web Attacks" in 2010 according to OWASP.

SQL injection is often caused by insufficient check for input. Therefore, high-quality code is the way against SQL injection. Literature [1] list some guidelines for safe programming and propose a method can find possible holes by code detection for web applications. Literature [2] proposed a new black-box testing method. For JDBC program, literature [3] proposed a method for detect static code. But only by improving the quality of the code or reduce code vulnerabilities can't solve all the problems, hackers can still find a new string of attacks to circumvent these program



checkpoint. Other protective measures also include injecting feature inspection. SQL injection feature filter means increase application firewall on the front end of web service, that can filter malicious input when receiving input information by centralized inspection, for instance, Snort and MagicQuotes which is implemented with PHP. Feature string matching algorithm that according to predefined regular expression is often used in SQL feature filtering program. On the one hand the performance will decreased rapidly if check lots of features, on the other hand the accuracy and false alarm rate is difficult to balance.

In this paper, a method about syntax tree feature-based matching approach to counter SQL Injection is proposed. On the one hand, this method can learn all legal SQL statement automatically and construct legal SQL statement knowledge base; on the other hand, this method also can identify legal SQL statement by pattern matching to achieve the purpose of protection.

There are three steps in this method: first, achieve SQL syntax tree by parsing all SQL statements under security circumstance; second, build knowledge base by feature extraction on SQL syntax tree; third, determine whether the legal SQL statement by pattern matching under actual circumstance.

The key points in this paper are as follows:

- 1) the implementation of SQL Injection protection using pattern matching;
- 2) pattern feature extraction based on SQL syntax tree is more accurate and more efficient;
- 3) generate all SQL statement automatically, build knowledge base;
- 4) use an efficient HASH algorithm, and this can improve the matching accuracy and efficiency in the process of pattern matching.

The content as following: section 2 and section 3 introduce the basic principles of SQL injection protection based feature matching, and the process; section 4 introduce the specific implementation of the prototype system; section 5 introduce the performance of test on prototype system function in some specific power business system; section 6 introduce the insufficient of this method, and the corresponding improvements.

## 2 The Principle

For specific application system, the number of isomorphic SQL statements can be controlled, and statement syntax will be changed if it contains any SQL injection statement, for instance, adding or deleting some node, change the structure of subtree. Therefore, if SQL statement syntax is inconsistent with the expectation, the statement can be considered as SQL injection statement. For the above mentioned premise, SQL injection protection base on syntax tree feature matching can be implemented. The method will be divided into two phases for SQL injection protection: learning phase and filtering phase. In learning phase we achieve SQL syntax tree by parsing all SQL statements under security circumstance, build legal SQL statement knowledge base. In filtering phase, we determine whether the statement is legal by pattern matching under actual circumstance. The main principle of the method involves the following:

### 2.1 SQL Syntax Tree and Feature Extraction

To obtain SQL syntax tree, we need parse all SQL statements through lexical, grammatical analysis. After traversing and trimming the tree, we will make it to be a standard syntax tree. During trimming process, user input, such as numbers and strings will be replaced with a wildcard, and some useless nodes will be removed. Then, we can extract feature of SQL statement based on standard SQL syntax tree. The features include number of child tree (t1), the root tree height (t2), number of tree nodes (t3), number of tree nodes in first child tree (t4), and type of first tree node (t5).

$$\varphi(i) = \{t1, t2, t3, \dots, tn\} \tag{1}$$

As shown in (1): n is the number of extracted features; t1, t2, t3, ..., tn are the eigenvalues;  $\varphi(i)$  is the eigenvectors of i-th standard SQL syntax tree.

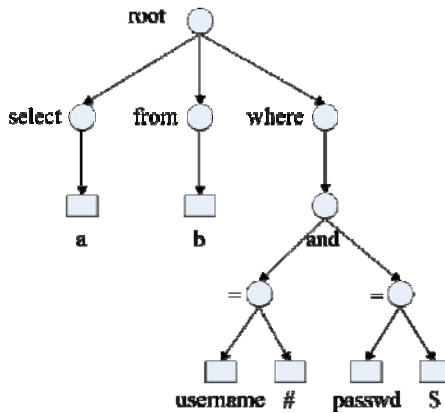


Fig. 1. SQL standardized syntax tree

Fig.1 is standard SQL syntax tree of the SQL statement “select a from b where username='mike' and passwd=123456”. After feature extraction, the eigenvector is  $\varphi(i)=\{3,4,13,2,1,\dots\}$ .

### 2.2 Knowledge Library and Pattern Matching

SQL knowledge library contains all different backbone structure of the SQL statements learned in learning phase, and is the basis for the real-time filtering phase. From an implementation standpoint, SQL knowledge library is memory pool which stores all SQL statements with different feature. In order to improve the efficiency of matching, internal storage of knowledge library is a static array with multiple - level index, as shown in fig 2. Pattern matching is based on structural features of the SQL statement syntax tree, as  $\varphi(i)$ . And then, get the eigenvalue (sql\_value(i)) using hash algorithm, like that:  $sql\_value(i) = Hash(\varphi(i))$ . Using the eigenvalue (sql\_value(i)) as an index, we can search knowledge library to find whether there is the same feature of SQL statement.

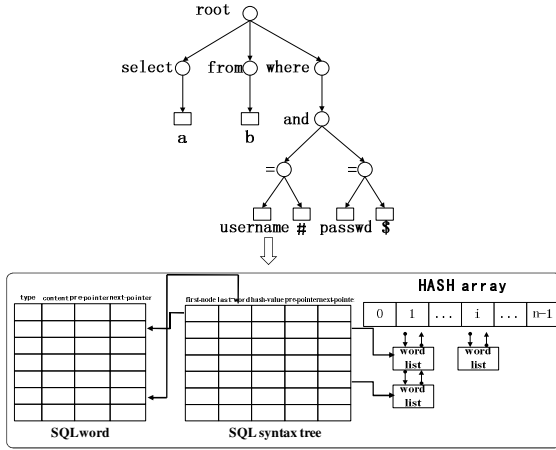


Fig. 2. Knowledge library storage

### 2.3 Value-At-Risk and Feature Filtering

Illegal SQL statements often contain some special characters or strings, such as empty string, 'or', 'union', the comment symbol, constant, and so on. However, not all SQL statements which contain these special characters or strings are illegal, so there is inevitably false alarm rate for feature filtering. In order to improve filtering accuracy and reduce false alarm rate, in this paper, we use a dynamic feature filtering method. by giving each character string a risk value. For example, the risk value of “empty string” is 3, and the risk value of “or” is 1. Then we define a threshold for SQL injection. One SQL statement is illegal only when all the accumulated risk value is greater than the predefined threshold. Calculation formula as below:

$$R = \sum_{i=1}^n r_i * k \tag{2}$$

R is the accumulative risk value of this SQL statement, n is the predefined maximum number of feature type,  $r_i$  is the risk value of i-th injection feature, k is the number of occurrences of an injection feature. By dynamically adjusting the risk value of various injection features and the total threshold, this feature filtering method is more flexible and has better accuracy.

## 3 Detailed Process

### 3.1 Learning Phase

As shown below in Fig 3:

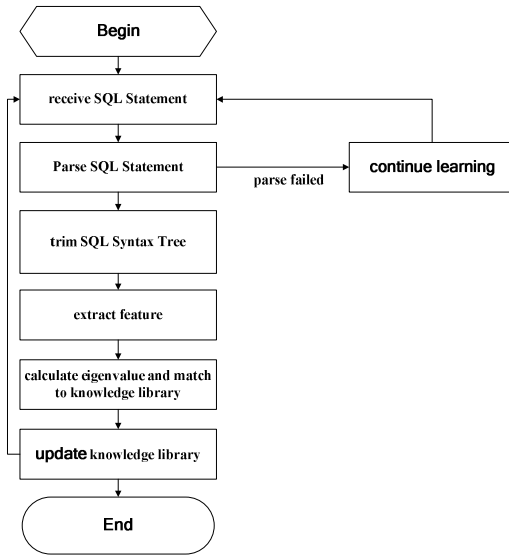


Fig. 3. Learning phase

### 3.2 Filtering Phase

As shown below in Fig 4:

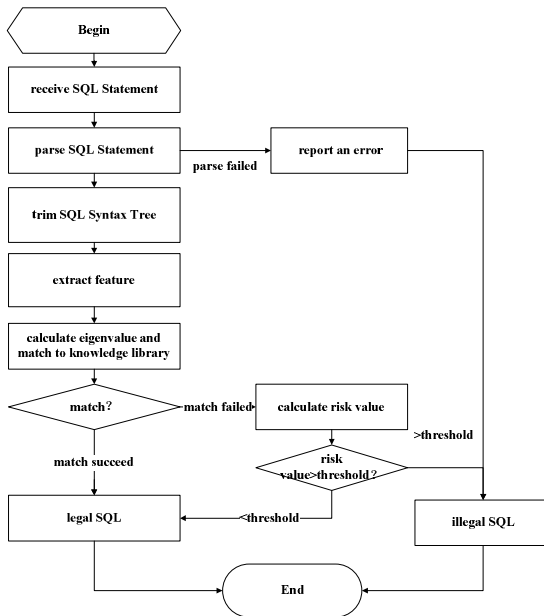


Fig. 4. Filtering phase

## 4 Design and Implementation

This section describes the design and implementation of prototype system, the module diagram as follows:

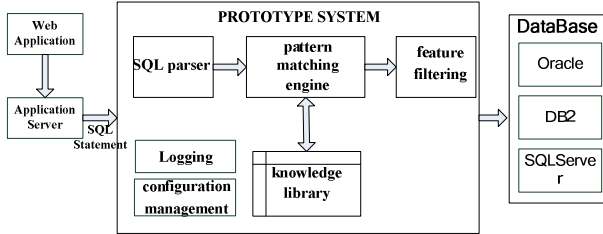


Fig. 5. Prototype system modules

**SQL parser:** responsible for parsing SQL statements including lexical analysis, syntax analysis and semantic analysis. Lexical analysis is responsible for decompose the original SQL statement stored in a continuous period of storage buffer to separate words which will be organized to a word list. Syntax analysis is responsible for constructing an SQL syntax tree based on the word list. Semantic analysis is responsible for achieving the semantic features by analyzing the SQL syntax tree.

**Pattern matching engine:** responsible for trimming the SQL syntax tree according to the scheduled policies including replacing user input, deleting useless nodes and so on. Then, compute eigenvalues for this SQL statement and search knowledge library.

**Knowledge library:** responsible for storing all SQL statements with different feature. In order to improve the efficiency of matching, internal storage of knowledge library is a static array with multiple - level index.

**Feature filtering module:** responsible for detailed feature checking, calculating risk value. Then, determine whether the SQL statement is illegal based on the threshold.

**Configuration management module:** responsible for receiving the local administrator or remote management client instruction, then dynamically update the local policy.

**Logging module:** responsible for receiving and logging the alarm information.

## 5 Modeling

To verify our approach for sql-injection detection, we use a simple modeling method. Suppose through full of learning, legal SQL statement coverage  $f(0% < f < 100%)$  in knowledge library. Feature filter algorithm based on value - at - risk has the accuracy rate of  $a(0% < a < 100%)$  for illegal statement, has the misinformation rate of  $b(0% < b < 100%)$  for legal statement. We define that the accuracy rate is A, and the misinformation rate is B by using our method.

In this example, we assume that the total number of SQL statements is T, legal statements T1, illegal statements T2. For legal SQL statement,  $T1 * f$  statements match knowledge library. And the remaining  $T - T1$  statements are judged to be

suspicious which will be checked by feature filtering module. Then  $(T1 - T1*f)*(1-b)$  statements will be judged legal, and  $T2*a$  statements illegal.

So, accuracy and misinformation rate for the entire module is:

$$A = \frac{T2*a}{T2} = a \tag{3}$$

$$B = \frac{(T1 - T1*f)*b}{T1} = (1 - f)b \tag{4}$$

As you can see, our approach mixing pattern matching and feature filtering does not directly enhance the accuracy rate for illegal SQL statements, but significantly reduces the misinformation rate for legal SQL statements. The coverage is higher for legal SQL statements in knowledge library, the effect is more significant. So, in order to improve the capacity for detection injected SQL statements, we can increase inspection during feature filtering while still able to maintain low misinformation rate. This can be very good to avoid contradictions between accuracy rate and misinformation rate.

## 6 General Analysis

To improve test accuracy, we select two real business systems, and set up simulation business environment. The prototype system is set between WEB server and database server as a proxy.

### 6.1 Performance

As the prototype system is an additional processing for SQL statements, it will partly increase the total time for user access to the database. We select 128, 256, 512, 1024, 2048-byte length SQL query statement. We execute each statement 10,000 times, and then get the total processing time. Test results shown in the following figure 6:

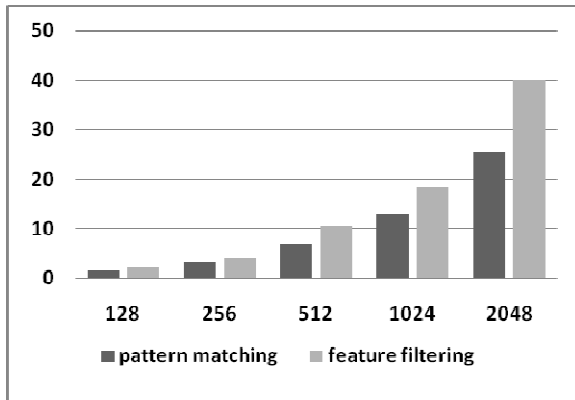


Fig. 6. Performance analysis

According to test results, the system processing time for SQL statements is directly related to the length of it. Pattern matching has greater advantage with feature filtering in processing time. According to test, the total processing time for general database queries (10 rows of data and 100,000 times) is 130 seconds. Therefore, the additional time for both pattern matching and feature filtering is far less than the time for general database access. And since most SQL statements only pass by pattern matching rather than feature filtering, the overall process performance will be higher.

## 6.2 Function

Functional test is to test the ability of the prototype system to detection SQL injection. First, we learn as much as possible all legal SQL statements of the two business system in learning phase. Through a period of learning, the total SQL statement in knowledge library of one business system is 1256, the other 469. Test results show:

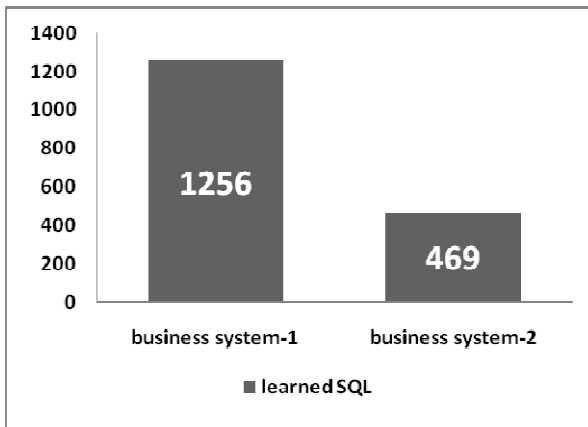
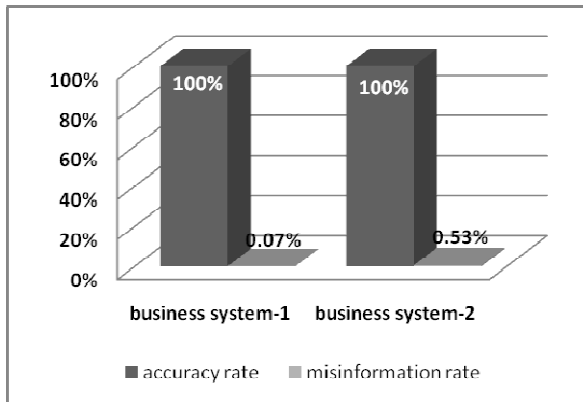


Fig. 7. Learned SQL

Then we construct 200 illegal SQL statements and 1500 legal SQL statements. For one business system, the prototype system identifies 201 illegal statements and 1499 legal statements. And the accuracy rate for illegal statements is 100%, the misinformation rate for legal statements is 0.07%. For the other business system, the prototype system identifies 208 illegal statements and 1492 legal statement. And the accuracy rate for illegal statements is 100%, the misinformation rate for legal statements is 0.53%.



**Fig. 8.** Accuracy and misinformation rate

## 7 Conclusions

This paper presents a new approach for sql-injection detection, mixing pattern matching and feature filtering, which can void the defects of a single method. This method is easy to implement for a variety of typical deployment of Web applications, but also has better scalability. With the increased complexity of business systems, some applications may generate variable SQL statements for different user input. This brings great challenge to our method. On the one hand it leads to knowledge library not to fully cover all legal SQL statements, on the other hand a large number of variable SQL statements results in the capacity of knowledge library is huge. All of these will affect the accuracy of the method. In the future work, our will use the similarity of SQL syntax tree for pattern matching to optimize the method. This may solve the problem of variable SQL statements.

## References

1. Howard, M., LeBlanc, D.: Writing Secure Code, 2nd edn. Microsoft Press, Redmond (2003)
2. Huang, Y., Huang, S., Lin, T., Tsai, C.: Web Application Security Assessment by Fault Injection and Behavior Monitoring. In: Proceedings of the 11th International World Wide Web Conference (WWW 2003) (May 2003)
3. Gould, C., Su, Z., Devanbu, P.: JDBC Checker: A Static Analysis Tool for SQL/JDBC Applications. In: Proceedings of the 26th International Conference on Software Engineering (ICSE 2004) – Formal Demos, pp. 697–698 (2004)
4. Maor, O., Shulman, A.: Sql injection signatures evasion. White paper (April 2004), [http://www.imperva.com/application\\_defense\\_center/white\\_papers/sql\\_injection\\_signatures\\_evasion.html](http://www.imperva.com/application_defense_center/white_papers/sql_injection_signatures_evasion.html)



5. Ashcraft, K., Engler, D.: Using programmer-written compiler extensions to catch security holes. In: Proceedings of the IEEE Symposium on Security and Privacy, May 2002, pp. 143–159 (2002)
6. Buehrer, G., Weide, B., Sivilotti, P.: Using Parse Tree Validation to Prevent SQL Injection Attacks. In: Proceedings of the 5th International Workshop on Software Engineering and Middleware, Lisbon, Portugal (September 2005)
7. Su, Z., Wassermann, G.: The Essence of Command Injection Attacks in Web Applications. In: Proceedings of the 33rd Symposium on Principles of Programming Languages, Charleston, South Carolina (January 2006)
8. Zhou, J.-L.: A New Policy to Defend against SQL Injection Attacks. *Computer Science* 133(111) (2006)
9. Yu, J.: Research on penetration testing based on SQL injection. *Computer Engineering and Design* 28(15) (2007)
10. Chen, X.-B.: Research on technique of SQL injection attacks and detection. *Computer Engineering and Applications* 43(11) (2007)

# Cloud Computing Concept in the Application of HIT Library

Xiaodan Wang, Tingbo Fu, Guang Yu, and Jiaomei Sun

Library, Harbin Institute of Technology, HIT, No.17 Siling Street, Nan Gang District,  
Harbin, 150001, China  
{wangxd, futb, yug, skali}@hit.edu.cn

**Abstract.** In this paper, our library based B/S model based on the library automation system, combined with cloud computing technology, analysis of cloud computing technology, application prospects in the library, as well as the library facing the management, data security, copyright, and other aspects of the challenges. We should take action to meet the new technological revolution has given us new opportunities and challenges.

**Keywords:** Cloud computing, Library, Technical standards, Data security.

## 1 Introduction

Cloud computing is a distributed computing technology, its basic concept, is an enormous amount of network computing automatically split into numerous smaller the subroutines were then referred to the composition of multiple servers large system, through search, calculation, and analysis ,will address the results back to the user. With this technology, network service providers can deal with billions of information within a few seconds, can achieve the “super computer” power. Cloud computing used in the network services. Such as search engines, internet mailboxes, users enter a simple command that can get a lot of information [1].

## 2 Cloud Computing

### Based on B/S Model Library Service System

Our old library automation library system is based on C/S mode system, in the LAN environment, the better to meet our library's basic business processes. However, with the continuous deepening of library work, such as real-time foreign mining operations, outsourcing of cataloging services, are all put forward a new automation system requirements. C/S mode on the shortcomings exposed. In order to meet library development under the new situation, our library and software company jointly developed the H-LIB system, the new system has based on B/S model. The new automation systems achieves through the WWW browser , little part of the business logic achieves in the Browser, the main business logic implements on the server side,

formats the so-called three-tier structure. This greatly simplifies the computer loads of client, reduces the system maintainable cost and the system upgraded effort.

### H-LIB System Circulating Interface

We use the H-LIB library automation system in the three campus now. To provide facilities for CALIS interlibrary loan between libraries, the various members of the library use the same H-LIB system. You can search OPAC between the library and the library of real borrowing and returning books.

Barcode	Title	Author	Index	Lend Date	Return	Renew	site	Price
4010850000	Impression of light Material / Rendering technique	Liu bo	TP391.41/L631	10-03-17 15:24:29	10-05-17	0	First Campus	58.00
4010190263	C# Network Programming	Liu ruolin	TP312C/L71-2	10-03-17 15:04:20	10-05-17	0	First Campus	34.00
4010388933	C# Advanced Programming	(U.S.)Simeen Hall	TP312C/L80	10-03-03 14:50:28	10-05-03	0	Second Campus	128.00
4010308029	V-Ray Raiders Super Render	(Korea)jin yingqi	TP391.41/091	10-02-27 08:47:45	10-04-30	0	First Campus	78.00
4010596189	V-Ray 1.5 Self-paced tutorial	Li bin	TP391.41/L191-2	10-01-19 15:05:47	10-04-30	1	First Campus	79.00

Fig. 1. H-LIB system circulating interface

### Currently Using a Database in Our Library

Currently our ChaoXing E-books and CNKI adopts the form of Database package, which greatly saves the running cost. We do not need to buy server and storage hardware, needn't install and maintain software, needn't maintain updated data resources. As a result, this great saves my area of space, reduces energy use, saves the expenses of server and storage hardware, reduces the workload of maintenance. It has taken into consideration requirements of environmental protection. This kind of package database uses the digital resources, much like the use of power resources now. Now we use power, just need to install the ammeter to statistics, pay according to ammeter reading, rather than building its own power plant. Currently most of the libraries access to the database adopting the from of database package.

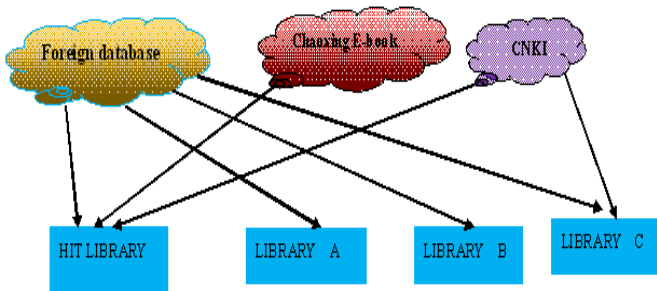


Fig. 2. HIT Library Database

### 3 Cloud Computing Applications

#### Cloud Computing Main Applications[2]

##### Amazon Website

Online bookstore and e-retailing is started, Amazon is now renowned in the industry, but it has the latest business and cloud computing related. About three years ago, as one of the first manufacturers in emerging markets, Amazon enter the cloud computing. It created a good start for other companies to enter the this field. Amazon called Amazon Web Services Cloud, it mainly composed by the four core services: simple storage services, elastic compute cloud, simple arrangement services, and simple database of testing phase. In other words, Amazon is now offering network access through network storage, computer processing, information queuing and database management system.

##### Google, Inc.

Google created a super powered business model around the internet search. Now, they used application hosting and enterprise search, mostly opened to businesses in the form of their "cloud." In April last year, Google pushed out Google Applications software engines (Google Apps Engine, GAE), this service allows developers to build on Python applications, and free uses Google's infrastructure to be host (maximum storage space up to 500MB). Exceed this limit for storage space, Google provides fee-based services, charges 10 to 12 cents per CPU core hour and 1GB space charges 15 to 18 cents.

##### Salesforce

It is a pioneer in Software-as-a-service, which started to provide network access is available through the sales automation applications. Under the leadership of company, other companies also prospered. Next Salesforce's goal is Platform-as-a-service.

##### Microsoft

In the initial stage of cloud computing, Microsoft has experienced many setbacks. After several years of running and adjustment, the software giant's cloud computing strategy was finally on the right track. According to some manufacturers expected, most of the IT resources of the future will come from cloud computing, but Microsoft did not think so. Microsoft's chief software architect (CSA) Ray. Ozzie has said that Microsoft's grand plan is to "provide a balanced mix of enterprise's software, partners' hosting service and cloud services." In short, Microsoft will be referred to as "software plus services."

#### Cloud Computing in the Library of Application Prospect

There is not long about the concept of cloud computing, but because of the background of cloud computing technology and the driving force is very complex, the researchers mixed background and vision, Thus, many of the concepts of cloud computing are defined. Matrix has summarized the 20 definitions of cloud computing, [3] but still far been covered.

The CNKI Database used the mirror image station of our library, the disc was mailed monthly and copied to the local storage devices. Now I have used CNKI Database in full package, but also keep the disc back in our library. Package database

did not submit applications in the form of the concept of cloud computing, but in fact is the library connected to the "CNKI database cloud."

Such as library automation management system, we don't purchase local servers and local storage, don't maintenance system software upgrades and database backup and so on. All of these can be handed over to the software company for installation and maintenance. We only need to pay the annual fee, use library automation management system in the form of rent. This can save a lot of hardware investments and labor costs. [4] If the software company's services can not meet our requirements, we also can change other software company choose a better company to meet the needs of our software system.

With Amazon, Google, Microsoft entering the "computing cloud war", our library can not stand by and from the cloud computing technologies, our library should explore and apply cloud technology, and grow together with it. Cloud computing applications in the library a bright future, I believe that it will give us great revolution and new challenges.

## 4 Cloud Computing Challenges to Library

Currently, libraries from application of cloud computing will be many difficulties, we must explore in the application, there is a greater risk of process than the previous library automation and digital library. [5] With the further development of cloud computing technology, we believe the concept of cloud computing has been widely accepted by users. Cloud computing will become the most important infrastructure. The application of cloud computing platform will become the most important choice by users. [6] So we have to consider the following questions:

### The Preservation of Library Resources in the Cloud Age

According to foreign media reports (Beijing Time October 12, 2009), Microsoft's Danger issued the statement, because the servers have faults, Sidekick mobile phone user's address book, calendar, photos and other personal data have been lost. The faults to make Microsoft and T-Mobile very embarrassing, but also is possible that the storage service of cloud computing has been potentially dangerous. [7]

In theory, the use of cloud computing technology, users only need a local terminal to access to the library-related resources through internet. But in fact, we consider data backup of great importance for the security of data on "cloud storage". Therefore, according to distinguish between the importance of library data, some data sources can be placed in the "cloud storage" and does not require backup; some data sources can be placed in the "cloud storage" to backup; some data must be placed in the of existing models to be retained. This is what our librarians should consider in the cloud computer age.

### Copyright Issues

In the cloud computer model, through the agreement between the libraries, share each other's information resources. The librarians have the same rights in use of ours and

other libraries. So this was potential copyright issues. Thus libraries must make a sharing information plan to make convenient access.

### **Construction of the Network**

Cloud computing is built on the network. Whether the library resources can be put to good use and sharing, security and stability of the network is the most basic and important.

## **5 Summary**

The technology and service of cloud computing have been applied to the library. In the beginning of the new technology, it will bring both positive and negative effects. The advent of cloud computing technology will be bringing huge challenge and opportunity for library development. We have met the challenge and triumphed.

### **Acknowledgments**

Name: wang xiaodan

Address: Library, Harbin Institute of Technology, HIT, No.17 Siling Street, Nan Gang District, Harbin, 150001, China

Email: wangxd@hit.edu.cn

### **References**

1. Cloud computing (April 10, 2010),  
<http://baike.baidu.com/view/1316082.htm>
2. Cloud computing main applications (April 10, 2010),  
[http://baike.baidu.com/view/1316082.htm?fr=ala0\\_1\\_1](http://baike.baidu.com/view/1316082.htm?fr=ala0_1_1)
3. Matrix.20 definition of cloud computing (April 11, 2010),  
<http://internet.solidot.org/article.pl?sid=08/07/18/0511233>
4. Hu, X., Fanbingsi.: Cloud computing presents challenges to the library management. *Journal of Academic Libraries* (4), 7–12 (2009)
5. Fan, B.: Cloud computing and Library: Defense for cloud computing research,  
[http://docs.google.com/View?id=dcmwg7xg\\_289hmbk35hj](http://docs.google.com/View?id=dcmwg7xg_289hmbk35hj)
6. Cloud computing and storage problems of digital resources (April 15, 2010),  
[http://www.topoint.com.cn/html/cloud/cloudcom/2010/04/282196\\_2.html](http://www.topoint.com.cn/html/cloud/cloudcom/2010/04/282196_2.html)
7. Cloud terrible harm has occurred, save the user data may be bankrupt (April 16, 2010),  
[http://blog.sina.com.cn/s/blog\\_59e64c8e0100fwmj.html](http://blog.sina.com.cn/s/blog_59e64c8e0100fwmj.html)

# Customer Segmentation and Marketing Strategy of Commercial Banks Based on CLV

Xia Wang and Lan Liu

School of Information Management Shanghai Finance University, Shang Hai, China  
Wangx@shfc.edu.cn, wangxia93@163.com, Liul@shfc.edu.cn

**Abstract.** Customer segmentation is one of the important research topics of customer relationship management. Through analysis of the information on retail customers, an evaluation indicators system is provided based on CLV. According to the customer segmentation of results, customer characteristics of each type and the corresponding customer marketing strategy are analyzed in detail, so as to provide support for the banks conducting objective evaluations and segmentations of retail customers' value, promoting optimal allocation of resources.

**Keywords:** retail customer, customer lifetime value, commercial bank, customer segmentation, marketing strategy.

## 1 Introduction

In the strategic transformation process of commercial banks operating, the retail business has become a major profit growth. However, domestic commercial banks are facing great challenges to keep existing customers and increase market competitiveness due to the huge retail banking market and the influx of foreign banks. Meanwhile in the financial industry, it is hard to distinguish financial services products and new financial products can easily be imitated. So effectively to customer segmentation, taking different strategies of sales and service, enhancing the level of customer satisfaction and loyalty, all of those are important ways to achieve diversity management of commercial banks in the trend of homogenization of financial products. And they are the important means to enhance the competitiveness of the market [1].

In this paper, a customer lifetime value (CLV) evaluation indicators system is provided, which is based on the based on the evaluation of the customer value system combining with the client attributes, behavior, risk factors and other non-monetary factors. On the basis of CLV as the key segmentation variables for customer segmentation, Strategic customers are found for the significant long-term development of commercial banks to guide the healthy development of customer relationships, and ultimately to achieve customer value maximization with the reasonable investment.

## 2 Construct the CLV Evaluation Indicators System of Commercial Banks

Customer Lifetime Value (CLV) refers to the total net present value of profit in the whole customer life cycle of each trading period [2]. Abroad, the theoretical calculation of CLV and practical application has matured. It is generally used to quantify, track, analyze customer value and predict the customer expectation value under certain assumptions. Based on them, bank's resources are reasonably allocated, and the long-term profit maximization is reached. Internationally famous banks such as Citigroup, America, HSBC, JP Morgan, are all practitioners and beneficiaries of effective customer management. In China, with the development of customer relationship management, China's commercial banks also began to focus on CLV in the retail business.

The process of calculation and assessment for CLV based on the customer life cycle is the comprehensive dynamic and is a dynamic process based on the time latitude. It is quite different from extensive management of the retail customer value which exists in China's commercial banks for a long time. The current value of the retail customer determines the bank's current profit level and helps finding the current high-value customer of banks. The potential value of the retail customer concerns with the bank's long-term profits and directly impacts on the customer value's subjective feelings and judge of banks in the next life-cycle. It is one of the important factors, which affects the banks continue to invest in the customer relationship. When the commercial banks are evaluating the value of retail customer, they are not only referring the current value of customers, but also predicting the potential value.

Thus, a multi-level evaluation indicators system is provided, which is based on CLV of the retail customer in commercial banks, shown in Table 1. The indicators system is not accurately used to calculate the value of a particular retail customer, but to do segmentation of the retail customers. It guides differentiated marketing strategy. Therefore, the simple and practical indicators system is also particularly important in addition to ensuring it is comprehensive and objective. In order to facilitate customer evaluation and selection for the bank, the indicators in this paper are mainly select from transaction data of retail customers in the banks as a source of research, which has good operability.

**Table 1.** The evaluation indicators system of retail customers in commercial banks based on CLV

First-level indicators	Second-level indicators	Third-level indicators
current value of retail customer	Assets business value X1	Total borrowed assets per day X11
		Average interest rate X12
	Liabilities business value X2	Total deposits per day X21
		Average interest rate X22



**Table 1.** (Continued)

	Intermediate business value X3	Average intermediate business income X31
		number of transactions X32
	Bank customer cost X4	Average operating costs X42
potential Value of retail customer	Loyalty Y1	Customers personal factors Y11
		Customer's Utilization rate of bank's productions and services Y12
		Changes in products or services Y13
		Customer Relationship duration Y14
	Credit rating Y2	

**The Evaluation Indicators of Retail Customer's Current Value**

China's commercial banks offer retail customers a wide range of banking. According to the form and function of different services, they are divided into three categories. The first is assets business such as personal housing mortgage loans, auto loans, other consumer loans, small loans, business credit card. The second is liabilities business such as savings deposits, debit card. The last is intermediate business such as collection and payment, escrow, agency, personal transfer, and consulting services, etc. [3]. All of above can bring benefits directly to the bank, which are used to evaluate the current value of retail customers.

The value of client assets is the proceeds brought by the customer's assets business, which the income is mainly interest income. Since the different loan species have different loan terms and interest rates, the bank's interest income will have various sizes. And there may be much assets business within the credit limits. So, the total borrowed assets per day and the average interest rate are used to evaluate as two indicators. Total borrowed assets per day =  $\Sigma$  the average balance of borrowed assets per day, average interest rate = total interest income of borrowed assets / total borrowed assets per day. Year period is used for the calculation.

The value of client liabilities mainly refers to the proceeds brought by the liabilities business to the bank. And different types of deposits have different interests. So the

total deposits per day and the average interest rate are used to evaluate as indicators. Total deposits per day =  $\Sigma$  the average deposit balances per day, average interest rate = deposit interest paid / total average deposits per day. And year period is used for the calculation too.

The value of client intermediate refers to the proceeds brought by the intermediate business to the bank. Since different intermediate business has different transaction costs or service fees and more transactions may bring more proceeds, so Average intermediate business income and number of transactions are used to evaluate as indicators. Average of the intermediary business income =  $\Sigma$  the intermediate business income  $\times$  the weight. The value of the weight is set by the commercial banks based on empirical data.

By analyzing the customer cost structure of commercial banks, customer cost includes the following four parts: acquisition cost, organization cost, transaction cost, and maintenance cost [4]. As in this paper the evaluation of customer value takes the transaction data as the main data sources, therefore, the computing of the customer cost is mainly taken the business into account. The average operating costs can be calculated by the cost rate and the client's business situation. In the management process the cost rate of each business type is determined by the bank based on business characteristics, marketing costs and other factors. Average operating costs =  $\Sigma$  the average balance of deposits per day / total deposits per day +  $\Sigma$  the average balance of loans per day / total borrowed assets per day  $\times$  the cost rate of loan. For the acquisition cost, maintenance cost and organization cost, some of them have been assigned to the banking business or products. Some of them are not considerations because these active marketing costs relate to the large amount of customers and these customers are often not identified.

### **The Evaluation Indicators of Retail Customer's Potential Value**

The basic idea of CRM is managing customers from the point of view of the customer life-cycle. So when the bank evaluates the value of the retail customer, it is not only to refer to the current value, but also the potential value which is not currently reflected in the income (profit) for the banks. This paper directly or indirectly evaluates the potential value of the retail customers from their loyalty and credit rating.

The behavior of loyal customers: incremental purchase, cross-purchase and recommending new customers [5]. The possibility of the incremental and cross-buying examines by analyzing customer's age, occupation, other personal factors and the use of bank products and services. The possibility of recommending new customers and influence need to take into account the state of relations between the customer and the bank such as the relationship duration, whether trusting banking products or services, customer status ( the occupation, age, job, etc). Therefore, the customer loyalty indicators are designed to customers' personal factors, the usage of banking products and services, changes in products or services, customer relationship duration.

Customers' personal factors include age, income, education level, occupation, family status. These components can be used five Likert scale indicators to measure and got a score representing the overall indicator value. Customer's Utilization rate of bank's productions and services = types of products or services in used / products and services provided by the bank. This indicator reflects the incremental purchase or the

possibility of cross-buying. Changes of products or services reflect customers' active state. If the state has not changed for a long time, the possibility of cross-buying and incremental buying is small. If the customers are active, such as new business during the investigation or the increasing trend of the transaction volume or the types from the performances of business history, the possibility is big. Therefore, this indicator can be measured by the growth rate of customer's business and business types.

Customer relationship duration reflects the degree of continuous usage of banking products or customer services. The longer that means the higher of the bank brand awareness and the satisfaction, the smaller possibility of customer churn, higher customer loyalty and the greater the likelihood of recommending new customers. The indicators can be in years, with five Likert scale calculations.

Credit rating is the bank's evaluation of customer credit. If a customer with low credit rating, the higher risk of bank credit, it is a large uncertainty to achieve the potential value for banks. A high degree of customer credit rating ensures that the expected benefits can be achieved and the realization of the potential value is stable and affordable. The indicator directly uses the value of bank's credit rating which has been obtained by the bank.

### **3 Retail Customer Segmentation and Marketing Strategies**

According to the above design of evaluation indicators system for the retail customer's current value and potential value, combining with relevant historical data, through clustering, neural networks, decision trees and other algorithms, the division based on these two dimensions can be achieved. The level of each dimension is divided into two tranches. Then, retail customers are divided into four groups, namely low-value customers, potential value customers, middle-value customers, high-value customers shown in Figure 1. Appropriate marketing strategies are developed for different customer segments.

I: low-value customers whose current and potential value are very low. For these customers, the bank needs adjustment or takes natural selection, without having to make too many resources. But the bank should keep its good reputation, a good current of each transaction to complete, at least not to become dissatisfied with the communicator.

II: potential value customers whose current value is low and potential value are high. They have a good trend to develop into valuable customers. For example, a customer's current unfavorable economic conditions or in terms of student customers whose age, status and income are changing over time. Such customers will bring more benefit to the banks. This type of customers needs to be developed and nurtured by the bank. The bank should carefully analyze the reasons why it fail to fully tap the potential value and the reasons why the loss of customers is easy. Implemented effective marketing strategy, they will bring higher returns in the future. The bank needs to enhance these customers' satisfaction and loyalty, and thus to fully tap its potential of contributing to the profits. The bank should give them more attention, so let them learn more about the business. And also the bank implements its marketing strategy of cross-selling, such as the provision of new services.

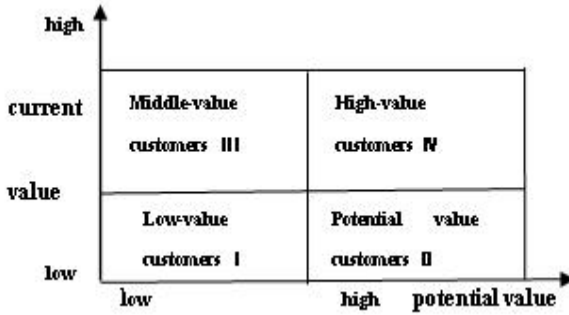


Fig. 1. Customer segmentation result based on CLV

III: Middle-value customers whose current value is higher but the potential value is low. And their space for later development is relatively small. Such customers' volume of current trading with the bank is larger and more satisfied with the provided services. But because of the limitations of the customer's own development, the possibility of the banks obtaining greater profits is small. The banks should maintain good relations with such customers, to prevent the loss of customers, because of their contribution to the current bank interest income as high, but without the need for developing cross-selling marketing plan and other resource inputs. These customers are instability, may leave at any time and turn to competitors. They should be implemented short-term strategy, and strived to maximize profits in the current period.

IV: High-value customers whose current value and potential value are high. They are the bank's high-quality customers. The relationship between these customers and the bank generally has access to the stabilization period. The bank has a larger share of such customer's current business volume. Such customers have great potential for future development and have many opportunities to bring more revenue to the bank. So, the bank should major resources into developing and maintaining relationships with such customers, fully understand customer needs and provide them with one to one service.

#### 4 Conclusion

The increasing emphasis on personal financial services business, the domestic commercial bank's daily operations are from the "product-centric" to "customer-centric". The retail customers in the banking business are becoming increasingly important. The traditional "Pareto's Law" (i.e. 80% of the profits of banks from 20% of key customers) has played a role in helping banks to understand and analyze customer. But it has some limitations. Scientific and comprehensive for customers segments become a top priority. CLV-based customer segmentation is very practical and effective method. In this paper, an evaluation indicators system is provided based on CLV and combined with the characteristics of the banking sector. With the upgrade of customer relationship management systems, retail customer's information will

gradually become more complete and accurate. In order to lay the foundation for quick launching of the retail customer segmentation, the evaluation indicators system should be adjusted in real time or change, making the system more accurate evaluation and data warehousing should be built strong and effective.

## References

1. Zhang, Z.-Y., Zou, R., Li, J.-F.: Identification of Customers' Satisfaction Index on Retail Business of Commercial Banks. *Finance Forum*, 36–41 (October 2006)
2. Qi, J.-Y., Shu, H.-Y.: *Customer Value's evaluation, modeling and decision making*. Beijing University of posts and telecommunications press (2005)
3. Li, J.: Personal Banking Services in China: Present Development and Future Trend. *Finance Forum*, 46–50 (October 2002)
4. Xu, X.-P., Li, X.-D.: The customer cost management studies of China's commercial banking. *Economic Tribune* 16, 93–94, 117 (2004)
5. Chen, M.-L.: Experimental research on determinants of customer loyalty. *Journal of Management Science in China* 6(5), 72–78 (2003)

# Design and Implementation of Media Independent Handover on Heterogeneous Wireless Networks

Chang-Yi Kao, Chin-Shyurng Fahn, and Jia-Liang Tsai

**Abstract.** Due to the great promotion of wireless network access technologies, users in different locations may be covered by different wireless access networks. It is more desirable to integrate heterogeneous access network technologies in multi-mode network devices for users. The IEEE 802.21 working group defined a framework to provide The Media Independent Handover Functions (MIH Functions) to achieve handover with better performance. However, most implementations of IEEE 802.21 were based on mobile IP. The cost of implementation of IEEE 802.21 is high. In this project, we design and implement the MIH Functions by low-cost approach within Wi-Fi and Bluetooth. We can provide seamless handover to users such that user can switch access networks within Wi-Fi and Bluetooth with low cost.

**Keywords:** wireless network, heterogeneous, Media Independent Handover (MIH), Wi-Fi, Bluetooth.

## 1 Introduction

As there are limitations on the range when using services provided on wired networks while no such restrictions, a need for a wire rod for instance, are imposed in an wireless environment, standards governing access techniques to wireless communications have been formulated and made known one after another. These access techniques include one series drawn up by the Institute of Electrical and Electronics Engineers, or IEEE for short, of the United States, which is devised with IEEE 802 as its prefix. It is also known as the IEEE family of specifications, such as IEEE 802.11 Wi-Fi (2) of the wireless LAN (WLAN) , IEEE 802.15.1 Bluetooth (1), IEEE 802.16 (3) of the Worldwide Interoperability for Microwave Access (WiMAX), and those not belonging to the IEEE family like the 3rd Generation Partnership Project, or 3GPP. Although the covered communications range and speed of transmission of the Net access techniques are different among themselves, users having mobile devices equipped with various network access interfaces are able to get network services from different sources and locations by use of different network techniques.

For mobile users to gain access to a variety of network services at all times and places, a multi-mode network device with a set of network interface cards have to be made available. For example, when an indoor wired IEEE 802.3 Ethernet is to be moved to the outside where services may be within the coverage of a wireless access point (AP) or a base station (BS), a switch from cables to wireless networks, the process of which is called “handover,” is necessary. If the handover occurs between

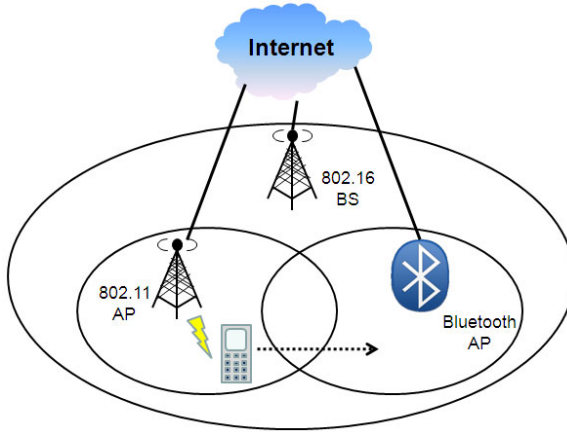
homogeneous networks, it is dubbed “horizontal handover.” For the switch taking place between heterogeneous networks, it is termed “vertical handover.” In the existing networking environment, the implementation of a multi-network access technique is becoming one of the focuses of the current network development. A mobile device is expected to be capable of getting access to various kinds of network services by use of the handover process for network access techniques supported by different interface cards.

## 2 Related Works

In [5], a Handoff Protocol for Integrated Networks, or PHIN, is proposed for integration of services through handover among different networks with a view to gaining a better performance for mobility node points on heterogeneous networks. The theory calls for an IP layer handover as the starting point to be based on the hierarchical mobile IP, aiming to make improvement during a massive exchange of information as a result of handover through the conventional mobile IP techniques. By taking advantage of the mobility anchor point (MAP) acting as proxy and through the definition of its messages, MAP is able to bear the load of each sub-network and control the link and mobile model of every node point on the subnet. As MAP is also capable of receiving in advance information surrounding the mobility node point and neighboring networks, it can facilitate the handover process making less the time for suspension on service during the switch, lowering the packet loss ratio, and maintaining the good quality of service. Nevertheless, system usually adopts a central management model which is not instrumental in the development of techniques for the multi-covered heterogeneous networks.

In [6] and [7], a reference is made to a media independent handover mechanism to be used for handling the handover techniques on heterogeneous networks. The mechanism discussed in [6] uses Mobile IPv6 as the basis to harness the movement of the mobility access point when exercising access to different networks. In [7], a handover process aimed at making the switch successful without much delay is proposed so that a seamless change from a wired environment to a wireless environment without sacrificing the quality of service is possible. Under this structure, a server is employed to manage the handover of the mobility node point, depending on the status of each network, the preference level of users, and the load balance of the networks to select the best possible target network. When dealing with how to integrate heterogeneous networks, what mentions in [6] and [7] is a media independent handover scheme based on the IEEE 802.21 standard to connect the switch on the second and third layers. But the fact is that the experiment and the assessment are done with a simulator. As far as [9] is concerned, it talks about a practice involving the wired 802.3 and the wireless IEEE 802.21 specifications and using Mobile IP as basis designed to solve the handover problem occurring during the access process on different networks. While the Mobile IP approach is made under an all All-IP environment, the establishment of a home agent is needed in order to speed up the update of the current site address of a mobile device. Because of the high costs for setup and the prolonged delay in link pending the assignment of a new address, this approach is not suitable for its application on commercial composite machines

(servers) and the mobile device. Therefore, the primary job is to provide a low-cost, speedy media independent switch mechanism allowing the mobile device handover between usable networks to maintain its high service quality during the switch process.



**Fig. 1.** An Integration Illustrating Mobile Device under Structure of Heterogeneous Networks

As Figure 1, this network structure integrates three wireless networks of different standards, namely IEEE 802.11, IEEE 802.15.1 Bluetooth, and IEEE 802.16, with each having its own speed of transmission, range of communications, and different communications quality. Let's assume that this mobile device originally goes through the wireless network access point of IEEE 802.11 for service, that the horizontal handover is made by switch from the current wireless network access point of IEEE 802.11 to the same access point of IEEE 802.11, and that the vertical handover is done by change from the original wireless network access point of IEEE 802.11 to the wireless network of IEEE.15.1 Bluetooth or to the base station of IEEE 802.16, then the coverage of service can be extensive and the speed of transmission faster, meeting the requirement of consecutive access to service.

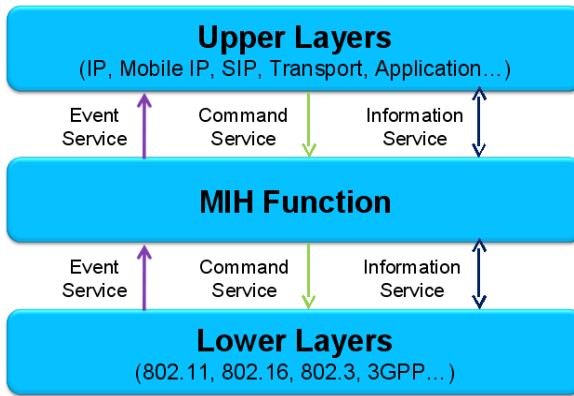
When handover occurs on the mobile device, be it horizontal or vertical, communications quality is unavoidably affected. To ensure a continued good and reliable communications quality, a seamless handover is what we need. Seamless handover means service is not stopped when a user encounters a situation where network sites change during the switch of networks for different access. In other words, the user does not get the feeling that the service is not successive when handover is being exercised.

### 3 Mobile Mih Function

The main purpose of this research is to put the IEEE 802.21 (4) Media Independent Handover, or MIH, into practice by integrating the two most popular wireless LANs,



that is, IEEE 802.11 Wi-Fi and IEEE 802.15.1 Bluetooth. Also, through providing the MIH mechanism cutting across the second- and third-layer handovers, the mobile device is able to gain a seamless link during the switch of access to different networks because the way of going through the second-layer handover leads to a less delay in time of switch and a reduced packet loss ratio. The following is a brief introduction to the IEEE 802.21 MIH scheme:



**Fig. 2.** The whole structure of MIH

The IEEE 802.21 MIH mechanism discussed here is aimed at introducing a scheme for integrating heterogeneous networks. MIH is primarily based on a cross layer offering a common interface for leaping over the second and third layers of networks. Networks at the upper layers need to register with MIH so that data at the lower layers, including the data link layer and the physical layer, can be collected, which are then transformed into a common mode for the upper layers to read. The upper layers, in accordance with the messages received, can issue orders to MIH at the lower layers for control of the operations on the networks at the lower layers. In the MIH mechanism, there are three types of services existing between the upper and lower layers, a device that governs the different network access media at the lower layers. The three services are Media Independent Event Service (MIES), Media Independent Command Service (MICS), and Media Independent Information Service (MIIS). Figure 2 is the whole structure of MIH, illustrating the interrelationships of the three services lying between the network layer and the data link layer in the MIH scheme.

The following is a detailed description of the three services.

### 1. Media Independent Event Service

In the general networking environment, the handover process usually takes place by going through the mobile device or the management of networks. The handover may be triggered by the mobile device or the network point of attachment (PoA) when sensing a status change on the network or in the management system at the network media access control layer or the physical layer. MIES describes a phenomenon showing that after networks at the upper layers have registered with MIH, they will inform their upper layers through MIH of the change of network

access media (including 802.3, 802.11, 802.16, and 3GPP2) at the data link layer. For example, in a wireless networking environment, when the mobile device completes the registration process and links to the access point, the condition at the lower layers of connection will be changed from linkdown to linkup. At this stage, MIH will touch off a report to the upper layers about the linkup allowing the upper layers to know the new status and take care of it accordingly. As the transmission pattern is different with different network access media, data from the lower layers must be transformed into a common pattern before being provided to the upper layers. To the upper layers, there is no need to be aware of detailed specifications of the access media at the lower layers. Thus, the media independent control can be achieved and a smooth handover in the heterogeneous networking environment brought about without difficulty.

## 2. Media Independent Command Service

MICS refers to a kind of service through which orders from the upper layers can reach the lower layers. Policy decisions from the upper layers may be transformed, through MIH, into network access media patterns at the lower layers, thereby networks at the lower layers can carry out comparable orders and allow the upper layers to keep an effective control of the different network access media at the lower layers as well as to make the superlative handover policies. For example, the lower layers, through MIH, make known an incident about to be off-line to the upper layers, indicating that now the mobility node point is likely to be leaving the access point previously connected. Provided that the lower layers have a number of network interfaces, then the upper layers can name a target network through the collected information and at the same time order the lower layers to switch to another network interface directing the information to a new network and restarting its services.

## 3. Media Independent Information Service

MIIS makes it possible for a mobile device during the process of moving to detect new access networks nearby and obtain useful information on the networks, paying the way for a smooth handover. As such, MIIS operates a query-response mechanism and goes through the media independent handover scheme to gain information concerning networks, including types of networks, channels of information, actual addresses of network access points, as well as the transmission speed supported by the access points, which are helpful for the upper layers to make decisions on handover. Therefore, the mobile device is able to use information obtained through MIIS to switch to the most suitable access network for a better service.

# 4 System Implement

When putting the Wi-Fi and Bluetooth handover systems into practice, this research finds that, as the starting link between Bluetooth and the server is faster, Bluetooth for communications is adopted at the very beginning. At the time Wi-Fi is linked to the server, switch is made from Bluetooth to Wi-Fi to continue receiving the residual packet for a better quality of service for users since Wi-Fi has a high-speed broadband. Under these circumstances, one needs to solve the problem how to let the server recognize that the two mobility node points, one being linked to the server through Bluetooth and the other switched to Wi-Fi at a later stage, are the same.

To achieve a seamless handover, the mobility node point needs only to register with the server once for all services it wants. In the practice, a user gets information first from the server through Bluetooth that is already linked to the server from the node point. At the junction of a successful link between Wi-Fi and the server, a notification is given of the fact that the node point is using both Bluetooth and Wi-Fi for communications and that the residual information will then be obtained from Wi-Fi.

1. System Component

In a heterogeneous networking environment, MIH has a cross layer structure used to integrate the switch process. For the upper layers, there is no need for them to know the media modes of the kernel layers. For the lower layers, however, a change must be made through the MIH mechanism to a corresponding mode in line with the packet for communications with the kernel layers. The function of service access points (SAPs) is to serve as a communications channel between MIH and other layers. Figure 3 describes the relationships between MIH and SAPs, consisting of two parts. The first part is about the upper layers and MIH\_SAPs while the second part concerns MIH and the MIH\_LINK\_SAPs of the network access media at the lower layers. Similar to a communications interface between the upper layers and MIH, access modes at the upper layers and the network access media are mutually independent; therefore, there is no need to switch to different modes because of different access networks. The main function of MIH\_SAP is to transform incidents occurring at the kernel layers detected through MIH to message modes compatible to those used in other communications agreements. As for MIH\_LINK\_SAP, it is an interdependent interface similar to MIH and the access media at the kernel layers. It must make a switch between the access modes provided by MIH and the different access networks.

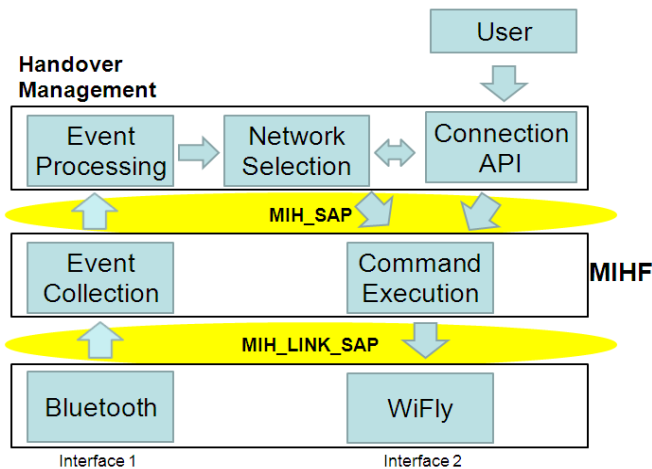


Fig. 3. The relationships between MIH and SAPs

2. Handover Process

Methods used in the research for the MIH system are illustrated in Figure 3.

- (1) First, the system will dispatch polling regularly to get the current link on the network.
- (2) When Interface 1 (Bluetooth) is on line, Event Collection will be notified that Interface 1 is in service.
- (3) Even Collection will call Event Processing for connection to the AP server.
- (4) Network Section now picks Interface 1 as the network interface, informing Command Execution that the current interface for the server is Interface 1.
- (5) User sends out Read Data Request to Connection API.
- (6) Connection API requests the server to read data the current network interface, that is, Interface 1.
- (7) Server will transmit back streaming data through the current network interface.
- (8) Interface 2 (Wi-Fi) at the client side notifies Event Collection of its being online.
- (9) Event Collection calls Event Processing to set up link (DHCP, IP, etc.)
- (10) Network Select now chooses Interface 2 the current network interface and asks Command Execution to notify the server through the current network interface that the current interface is Interface 2.
- (11) Server sends the residual data from Interface 2 to the client side.
- (12) If a break between Wi-Fi and the server or a bad connection occurs, Bluetooth can take over the communications. The practice is for Network Select to switch the network interface to Interface 1 and calls Command Execution to notify the server through the current network interface that the current interface is Interface 1. After that, Server continues to send the residual data from Interface 1 to the client side, completing a seamless service during the switch process.

## 5 Conclusion

This research attests to the practice that network access techniques are interchangeable with the mobile device. Currently, IEEE 802.21 draws up a draft examining the operation between Media Access Control Layers (MAC Layers) and Network Layers, aiming to allow the mobile device and the mobile base stations with different access networks all have an identical handover message, which can be put across for the mobile device to select various access networks for a smooth MIH process so that service is to be continued with a better quality during the transition on the heterogeneous networks.

**Acknowledgments.** This study is conducted under the "Project Digital Convergence Service Open Platform" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

## References

1. IEEE Std 802.15.1-2005, Standard for technology-Telecommunications and information exchange between systems-local and metropolitan area networks-Specific requirements-Part 15.1: Wireless medium access control (MAC) and physical (PHY) specifications for wireless personal area network
2. ANSI/IEEE Std. 802.11, Part 11: Wireless LAN Medium Access Control(MAC) and Physical Layer (PHY) specifications (September 1999)

3. IEEE Std 802.16-2004 TM, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems (October 2004)
4. IEEE P802.21TM /D00.05, Draft IEEE Standard for Local and Metropolitan Area networks: Media Independent handover Services (January 2006)
5. Makaya, C., Pierre, S.: Efficient handoff scheme for heterogeneous IPv6-based wireless networks. In: IEEE Wireless Communications and Networking Conference, March 2007, pp. 3256–3261 (2007)
6. Yoon, Y.A., Byung, H.Y., Kang, W.L., You, Z.C., Woo, Y.J.: Reduction of handover latency using MIH services in MIPv6. In: International Conference on Advanced Information Networking and Applications, vol. 2, pp. 229–234 (2006)
7. Yang, O.S., Choi, S.G., Choi, J.K., Park, J.S., Kim, H.J.: A handover framework for seamless service support between wired and wireless networks. In: International Conference on Advanced Computing Technologies, vol. 3, pp. 1–6 (February 2006)
8. Perkins, C.E.: IP mobility support for IPv4. RFC 3344 Internet Engineering Task Force (August 2002)
9. Shiau, Y.L., Jan, R.H.: Implementation of Media Independent Handover Functions for Wifi/Ethernet Heterogeneous Networks. MS Thesis, National Chiao-Tung University (2008)

# On-Line Detection and Evaluation of Cavitation in Large Kaplan Turbines Based on Sound Wave

Huixuan Shi<sup>1</sup>, Xuezheng Chu<sup>2</sup>, and Zhaohui Li<sup>3</sup>

<sup>1</sup> Wuhan NARI Limited Company of State Grid Electric Power Research Institute, China

<sup>2</sup> WISDRI Engineering & Research Incorporation Limited, China

<sup>3</sup> Huazhong University of Science and Technology, China

{shi\_huixuan, Chuxuezheng007, prof1zh}@gmail.com

**Abstract.** The work is dedicated to the development of an on-line monitoring and analysis system of cavitation in large Kaplan turbines (TrbMAU) in order to evaluate the cavitation degree in real time. Sound wave emitted by cavitation is continuously monitored, including audible sound and ultrasound.

Considering the influence of the operating states of turbine-generator sets on cavitation, adaptive data acquisition (DAQ) and storage is proposed. The DAQ period and storage vary with operating states to capture all sound features in different operating states with less data redundancy.

Based on the real-time evaluation of the signal characteristics, such as standard deviation, noise level, and frequency compositions, the tendency of cavitation intensity with time and different operating states has been traced out. Furthermore, the integrated cavitation intensity will be estimated periodically, which can figure out the degree of cavitation erosion approximately. And the research methodology and pivotal concerns are discussed on the evaluation of the metal loss caused by cavitation.

The TrbMAU has been successfully put into service in Gezhouba Hydro Power Plant. Its performance has been proved to be very good.

**Keywords:** Adaptive DAQ, Cavitation intensity, Erosion level, Kaplan Turbines, On-line, Selective Storage, Sound Wave, Tendency analysis.

## 1 Introduction

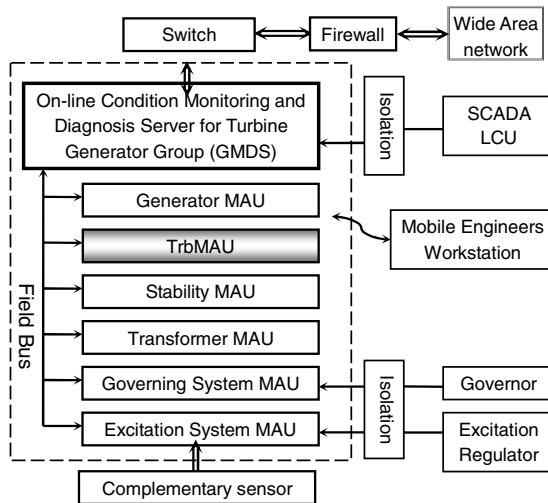
Cavitation erodes equipments, reduces efficiency as well as increases noise and vibration of hydro turbines. Today, about five turbines in Gezhouba Hydro Power Plant with 21 installations have to be repaired every year due to cavitation erosion. The maintenance decision is made mainly according to the service time. This is usually unreasonable. So it is highly desirable to have an on-line monitoring system which can indicate the cavitation condition in turbines timely.

To develop an on-line cavitation monitoring system, there are five key considerations:

- Detection approach: The feasible approach to detecting cavitation should have not or less impact on the security of turbine operation;
- Data acquisition (DAQ): The cavitation phenomena are featured by the operating states of turbine-generator sets. Therefore, the DAQ should be associated with the operating states so as to capture all sound features;
- Data management: The data quantity in an on-line monitoring system will be very large for long-term operation in the field, so an appropriate storage mechanism is desired to decrease data redundancy and release off-line analysis work.
- Sensors' placement: The location of cavitation is dependent on the types of hydro turbines. The source of cavitation, the features of sensors, the convenience and reliability of mounting sensors, the safety problems of turbine operation as well as the stability of flow, etc. should be taken into account in the sensors' placement;
- Cavitation evaluation: It is essential for condition-based maintenance of hydro turbines. In practice, the influence factors of cavitation, the features of cavitation erosion and the deterioration of turbines will be considered comprehensively.

So a Monitoring and Analysis Unit for Turbines (TrbMAU) has been developed. The sound wave emitted by cavitation is monitored to get insight into cavitation in turbines. The sensors' placement is discussed aiming at the large Kaplan turbines in Gezhouba Hydro Power Plant based on the nondestructive technique. Besides, the adaptive DAQ method and the selective storage mechanism are applied in TrbMAU. In particular, cavitation condition can be evaluated in terms of monitoring results in the field.

TrbMAU has been successfully serving in Gezhouba Hydro Power Plant as a component of the Optimal Maintenance Information System (HOMIS) (fig.1).



MAU: Monitoring and Analysis Unit

**Fig. 1.** Optimal Maintenance Information System

## 2 Detection Techniques

### Detection Approach

Presently, there are some methods for cavitation detection, such as stroboscope-aided visual observation [1], high speed photographing [2-3], measurement of pressure pulsations [4], as well as monitoring of sound pressure [5], cavitation ultrasound [6] or mechanical vibrations [7-8] on the turbine casing. The first two approaches are only used in tests. The others can be employed in field. In engineering application, the usage of pressure sensor and hydrophone in cavitation detection usually brings inconvenience and unreliability of installation as well as safe problems of turbine operation. Now, the preferred way of condition monitoring is based on nondestructive technique.

It is well known that sound wave will be emitted during cavitation, which included audible sound and ultrasound. The sound wave intensity can reflect the development of cavitation directly. Consequently, sound wave detection is greatly appropriate to get insight into cavitation in hydro turbines.

The generation, development, collapse, and rebound of cavities in water are always accompanied with ultrasound which predominates especially in the identification of cavitation inception [6]. Accordingly, acoustic emission (AE) sensors are employed to monitor cavitation ultrasound. Moreover, the audible sound will be emitted along with the development of cavitation, which is different from that resulted from flow [9]. Ambient in hydropower plants is quite noisy. So accelerometers are more suitable to collect audible sound or vibro-acoustic signals for their fine anti-jamming features.

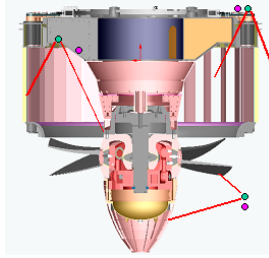
The proposed technique is an integrated acoustic solution. It can figure out the characteristics of both audible sound and ultrasound emitted by cavitation in real-time. And the solution has been comprehensively tested on the model turbine and the real ones. The test results demonstrate that it is suitable for on-line detection of cavitation in hydro turbines.

### Sensors' Placement

Sensor's placement is crucial for cavitation monitoring. The location of cavitation is closely related to the types of turbines. Usually, the nearer the sensor is to the source of cavitation, the more credible the monitoring results is. In application, the sensors are positioned with respect to large Kaplan turbines in Gezhouba Hydro Power Plant.

In TrbMAU, four accelerometers and four AE sensors are employed. And one accelerometer and one AE sensor are positioned in couples so as to broaden the monitoring bandwidth. They are mounted on the dry side of the turbine with the convenient magnetic assemblies based on the nondestructive technique. The locations of mounting sensors include guide van shafts, +X inner head cover, +Y inner head cover and the manhole wall of the draft tube. The sketch map of positioning sensors is shown in fig.2.





**Fig. 2.** Sensor positioning. Pink points and green points individually represent accelerometers and AE sensors. The acceptance angle of AE sensor is sketched by two red lines originating from each green point.

In sensors' placement, the following should deserve enough attention:

-- Cavitation source: there are four types of cavitation according to the place where cavitation occurs, airfoil cavitation occurring at the back of runner blades, clearance cavitation in the gap or clearance between the blades tip and the casing due to the local depression or the increase of flow velocity, cavity cavitation generated by unsymmetrical vacuum vortex at the runner trailing edge and the inlet of draft tube for off-design operation of hydro turbines, and local cavitation resulted from the uneven flow passage surface. For Kaplan turbines, an exclusive type of cavitation is the clearance cavitation. The cavitation erosion happens on the outside edge of blades [17].

-- Reliability and convenience of mounting sensors: non-destruct, reliability and convenience are significant in engineering application.

-- Flow stability in different places in hydro turbines. The instability of water flow is an important factor to induce cavitation. The inlet of spiral casing lies in +Y direction, where flow is very instable. While +X direction is responding to the tail end of spiral casing where flow is more stable.

-- Acceptance angle of AE sensors: The intensive directivity of ultrasound, the source of cavitation and the structure of hydro turbines are crucial to position AE sensors. Ultrasound attenuation in transmission should be considered, too. For example, it is much more in air than in water.

### 3 Instrumentation

TrbMAU has been stably operating as a part of HOMIS on 6 hydro turbines in Gezhouba Hydro Power Plant.

#### Instrumentation Structure

The operating states of the turbine-generator set and time are synchronized in HOMIS so as to comprehensively estimate the health of the turbine-generator set. The communication among MAUs and GMDS is established via CAN bus. GMDS and Governor MAU broadcast time information and operating states information in CAN bus in real time, respectively.

Analysis results from each MAU in HOMIS will get together and be displayed in GDMS to present health condition of the turbine-generator set to inspectors in the power plant. At the same time, comprehensive analysis and estimation about the system condition can be performed.

The structure of TrbMAU is illustrated in Fig.3.

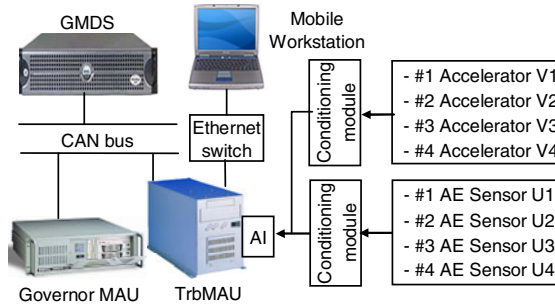


Fig. 3. On-line cavitation monitoring and analysis unit for hydro turbines

In addition, a mobile workstation for TrbMAU is designed based on TCP/IP protocol. It is not only an off-line analysis platform, but also is an operating platform in remote for TrbMAU, for example, data sampling and parameters setting can be performed according to the needs of field tests.

### Adaptive DAQ

The operating states of turbine-generator sets can be classified into transient states and steady states, for example, unit start-up, large load rejection and unit shutdown can rank transient states. They strongly impact the probability and intensity of cavitation. Obviously, it is preferred that the DAQ period varies with the operating states in order to ensure the integrality of data, which can be called as adaptive DAQ.

High-speed sampling and real-time analyses are involved in TrbMAU. The scheme is performed via DMA. Moreover, all analog input channels are simultaneous and synchronous. To meet the requests of on-line detection and tests, three DAQ methods, timing sampling, operating states triggering and manual triggering, are employed in TrbMAU.

For the transient operating states, continuous DAQ is used. Otherwise, the system will automatically switch to timing sampling to collect data. Meanwhile, a manual DAQ is available to pick up cavitation sound wave if necessary.

It is helpful for the on-line monitoring system that the adaptive DAQ method not only ensures to capture all sound features with less data redundancy, but also lessens the oppressiveness of system and increases the reliability.

### Data Management

Only cavitation signals are associated with operating states, could they be meaningful for cavitation evaluation in turbines [9-11]. In the sense, data storage will be based on

the operating states. Data stored include raw data, power spectrum, transient process, and detailed data. Raw data and power spectrum are stored in couples, and transient process is extracted via raw data analysis in transient operating states, while detailed data is from raw data analysis in steady operating states.

The selective storage mechanism can decrease data redundancy effectively. Different types of data are stored in different operating states. For transient operating states, only transient processes will be saved. While the other types of data may be stored only in steady operating states. Among them, the storage of raw data references power output, water head and noise level of cavitation. To be brief, raw data will be stored when the following conditions are fulfilled:

$$P_c - P_h > \alpha \times P_r$$

$$H_c - H_h > \beta H_s$$

$$NLC_i - NLH_i > NLS_i \quad i = 1, 2, \dots, 8$$

Where  $P_c$  - Current power output;

$P_h$  - Intraday power output in database;

$H_c$  - Current water head;

$H_h$  - Intraday water head in database;

$H_s$  - Design head of turbine-generator set;

$P_r$  - Rated power;

$\alpha, \beta$  - Coefficients, they can be modified;

$NLC_i$  - Current noise level of input channel  $i$ ;

$NLH_i$  - Intraday noise level of input channel  $i$  in database;

$NLS_i$  - Threshold specified for input channel  $i$ .

## 4 Cavitation Evaluation

TrbMAU has been successfully put into service on 3F, 5F, 10F, 14F, 15F and 19F turbine in Gezhouba Hydro Power Plant. And the 19F turbine is the first one to be monitored via TrbMAU in real time. TrbMAU has been serving on 19F, 3F, 14F turbine for more than two years.

### Cavitation Estimation

The cavitation estimation includes cavitation intensity and cavitation erosion level. It will be performed in association with the operating states.

(a) Grid of operating states

The operating states of turbine-generator sets can be denoted by power output, water head, and suction height. The energy level just upstream of turbines in Gezhouba Hydro Power Plant varies slightly in a whole year for the co-scheduling of Gezhouba dam and Three Gorges dam. So the water head of the turbines can represent the suction height approximately. That is, the operating states will be a planar expression of power output and water head.

The parameters of operating states vary significantly between day and night, as well as among seasons. To estimate the cavitation in turbines, the grid-based operating states are employed. Therefore, an operating state  $(P_i, H_j)$  can be represented with a grid. Usually, the interval of water head is no less than 5% times the design head, and the interval of power output is no less than 2% times the rated power [12].

In Gezhouba Hydro Power Plant, the water head is in the range of 8~27m and the design head is 18.6m. Thus, the grid of operating states of the turbine can be reported as fig.4.

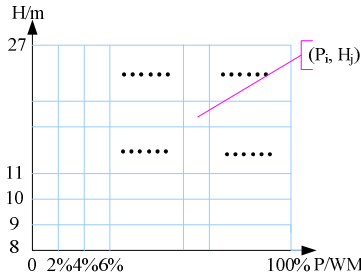


Fig. 4. Grid of operating states of 3F turbine

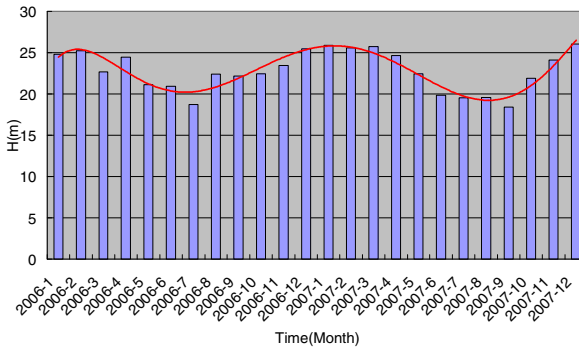


Fig. 5. Water head tendency of hydro turbines in Gezhouba Hydro Power Plant

Therefore, the probability of each operating state  $(P_i, H_j)$ , the average water head, and the average power will be achieved every half year according to the water head tendency in Gezhouba hydro Power Plant (fig.5), which illustrate the distribute of operating states of the turbine-generator set and can present evidences for the cavitation analysis and the optimization of turbine operation.

(b) Cavitation intensity

-- Real-time estimation: it is carried out mainly aiming at the load change tests which will be identified automatically;

Cavitation intensity  $I_{n(i, j)}$  in the  $n^{th}$  period of DAQ-analysis occurs at an operating state  $(P_i, H_j)$  which is different from that in the  $(n-1)^{th}$  period and becomes a new operating state until the  $(n+k)^{th}$  period. Then we can express the cavitation intensity  $I_{(i, j)}$  at the operating states  $(P_i, H_j)$ :

$$I_{(i, j)} = \frac{1}{K+1} \sum_{m=0}^k I_{(n+m)(i, j)}$$

-- Periodic estimation: the cavitation intensity under different operating states is reported weekly and monthly based on the daily analysis. And the cavitation intensity tendencies at fixed operating states are estimated annually based on the monthly analysis. In addition, the reports of the cavitation intensity tendencies in the time of more than one year can be requested manually.

The daily intensity  $I_{d(i, j)}$  of cavitation at an operating state  $(P_i, H_j)$  can be expressed as:

$$I_{d(i, j)} = \frac{1}{D} \sum_{n=1}^D I_{n(i, j)}$$

where  $D$  is the number of DAQ-analysis periods in the current day at the operating state  $(P_i, H_j)$ .

Then the monthly intensity  $I_{m(i, j)}$  of cavitation at the operating state  $(P_i, H_j)$  can be achieved:

$$I_{m(i, j)} = \frac{1}{M} \sum_{n=1}^M I_{d(i, j)}$$

where  $M$  is the days with the operating state  $(P_i, H_j)$  in the current month.

Besides, the integrated intensity of cavitation  $I_{intg}$  will be estimated every half year based on the water head tendency (Fig.5), the slow process feature of cavitation erosion, and the influences of different operating states on cavitation [9-11].

Based on the monthly intensity of cavitation, we have:

$$I_{hy(i, j)} = \frac{1}{H} \sum_{Half\ year} I_{m(i, j)}$$

where  $H$  is the months with the operating state  $(P_i, H_j)$  in the half year. Then

$$I_{intg} = \frac{1}{Y} \sum \sum I_{hy(i, j)}$$

where  $I_{hy(i, j)}$  is the half-yearly intensity of cavitation;  $Y$  is the number of different operating states (fig.4) in the half year.

In TrbMAU, the audible sound and ultrasound of cavitation are monitored separately and simultaneously. Moreover, the characteristics of them vary with the cavitation condition. Accordingly, it is essential for the cavitation evaluation to integrate the two types of signals. Different coefficients,  $\lambda_1$  and  $\lambda_2$ , are introduced according to the different attenuation of audible sound and ultrasound in transmission, and  $\lambda_1 \geq \lambda_2$ . The cavitation intensity  $I_{CL}$  can be defined as:

$$I_{CL} = \lambda_1 I_{intgU} + \lambda_2 I_{intgV}$$

where  $I_{intgU}$  and  $I_{intgV}$  indicate the integrated intensity of ultrasound and audible sound respectively.

### (c) Cavitation erosion level

It is to evaluate the metal loss resulted from cavitation via sound wave detection. The relative researches [12-13] have been made in the field. The method proposed in the document [12] is with the help of the cavitation intensity coefficient, which ignored the nonlinearity between cavitation erosion and serving time of turbines as well as the influence of low load on cavitation. While the literature [13] evaluates cavitation erosion level based on cavitation ultrasound, which is failed in serious erosion on the turbine blades. To improve the evaluation of cavitation erosion based on the nondestructive technique, we should pay enough attention to:

-- Signal transmission: the sound wave intensity of cavitation may decrease when cavitation develop up to some degree. The rupture of the liquid continuum will increase with the development of cavitation. Consequently, sound waves will be reflected or refracted in higher frequency at the boundary between air and liquid. In addition, signals attenuate more in air than in water, and the signal with higher frequency attenuates more seriously in transmission;

-- The sound wave emitted by cavitation will transfer from the high frequency to the low one along with the development of cavitation [14-15]. Thus audible sound and ultrasound caused by cavitation should be integrated to evaluate cavitation erosion;

-- The serving time of turbines with different maintenance history have different influence on cavitation [16];

-- It is quite important for the evaluation of cavitation erosion to accumulate the relative knowledge. And it is a good way to develop the comparative tests before and after runner blades' repair or renewal.

## **Cavitation Intensity under Different Operating States**

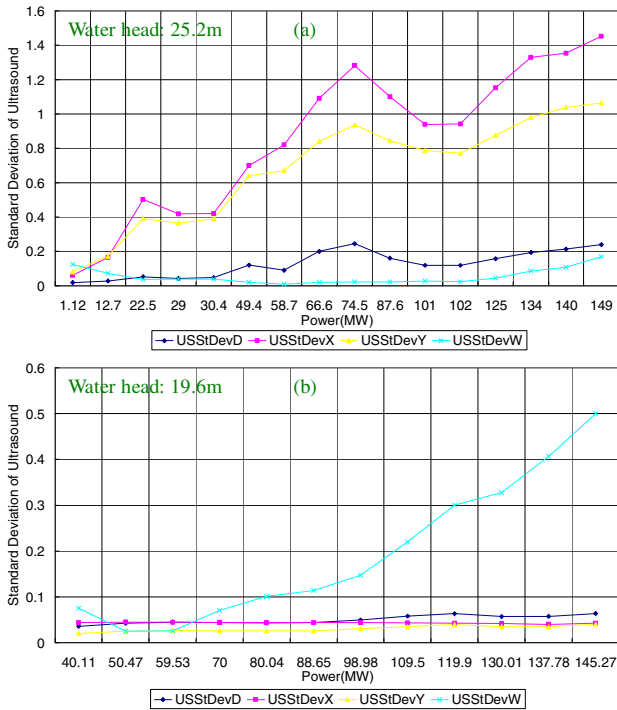
Load change tests on 3F turbine in Gezhouba Hydro Power Plant with 25.2m and 19.6m water head are reported in fig.6.

After comparing the two tendencies in fig.6 (a) and fig.6 (b), it can be found that the cavitation ultrasound level goes up gradually along with load increase. As to the cavitation ultrasound acquired at inner head cover and guide vane shaft, cavitation ultrasound level in 25.2m water head is more intensive than that in 19.6m water head. And cavitation ultrasound in 19.6m water head, heard at manhole wall from runner chamber, rises rapidly and is more intensive than that in 25.2m water head. However, cavitation ultrasound obtained at the other locations in 19.6m water head vary slightly and are much weaker. Apparently, 19.6m is very close to the design head (18.6m) of 3F turbine. It can be deduced that the sudden increase of ultrasound level achieved at the manhole wall in 19.6m water head may be resulted from the unsteady flow caused by large discharge.

From April to July, in 2006, water head varied from 25.2m to 14.9m. Meanwhile, the on-line monitoring results from TrbMAU installed on 3F turbine are strongly appropriate to evaluate the effect of water head on cavitation. The tendency of cavitation ultrasound along with water head in a static load on 3F is traced out, as shown in fig.7.

The ultrasound level of cavitation increases quickly in water head higher than 21.9m. And the cavitation ultrasound heard at manhole wall is much intensive in water head of 18.9m~20.9m. It is, however, much weaker and little varies in the other water head lower than 21.9m. Furthermore, the cavitation ultrasound, obtained at inner head cover and guide vane shaft, are much weaker and almost changeless in water head lower than 21.9m.

The phenomenon in water head of 18.9m~20.9m is the same to that in 19.6m water head in load change tests. Moreover, the passage frequency of runner blades in the raw data, heard at the manhole wall, can be easily identified in the water head range. According to the results of model tests, the more intensive the cavitation is, the more indistinct the passage period of runner blades is in the raw data [17]. Therefore, the phenomenon cannot be caused by cavitation, which is consistent with the facts of the hydro turbine.



**Fig. 6.** Tendency of ultrasound level along with load in different water head on 3F, USStDevD, USStDevX, USStDevY and USStDevW individually represent standard deviation of ultrasound obtained at guide vane, +X inner head cover, +Y inner head cover and manhole wall. Curves in fig.(a) and fig.(b) are separately obtained with water head of 25.2m and 19.6m.

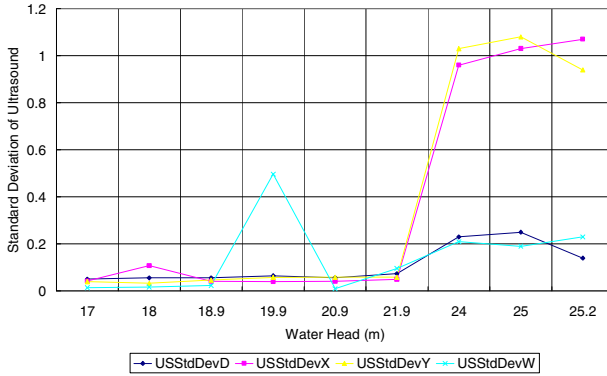


Fig. 7. Tendency of ultrasound level along with water head on 3F, P=146MW

The operating states with more intensive cavitation can be distinguished via the above analyses. For turbines in Gezhouba Hydro Power Plant, the higher the water head is, the more intensive the cavitation in turbines is. Therefore, these analysis results can be associated with the statistics of operating states, which will be great convictive for the prediction of turbine life and the optimization of turbine operation.

**Cavitation Intensity Tendency**

The grade A maintenance on 19F turbine in Gezhouba Hydro Power Plant was accomplished in 2003. And TrbMAU has been serving on the turbine since October, 2005. With P=100MW, H=26m, the tendency of cavitation ultrasound along with time is shown in fig.8.

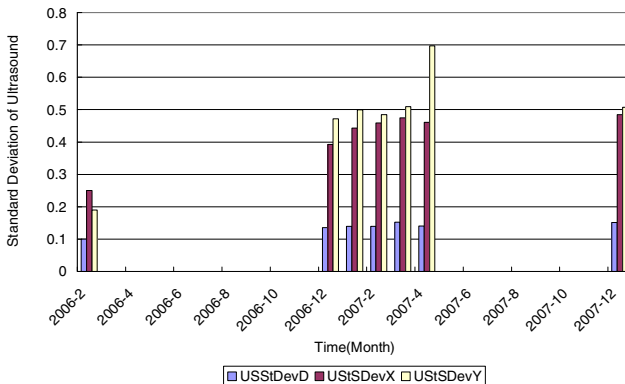


Fig. 8. Tendency of ultrasound level along with time on 19F, P=100MW, H=26m

We can see that the cavitation intensity is increasing gradually, and that cavitation is especially intensive in April, 2007. After comprehensive analysis, the unexpected



increase was found to be resulted from instable flow which was caused by the serious hunting of guide van servomotor.

Cavitation erosion is a slow process. Thus, it will be quite helpful to perform the tendency analysis of cavitation intensity for mastering the law of cavitation erosion along with the service time of turbines. Based on this, the condition-based maintenance of turbines will be conducted.

## 5 Conclusions

Cavitation is a common problem in hydro turbines, which causes serious erosion, reduces the performance of hydro turbines. On the requests of condition-based maintenance, TrbMAU has been developed. It can implement as an important part of HOMIS or operate independently.

Cavitation condition is achieved based on sound wave detection in TrbMAU. Both audible sound and ultrasound are monitored continuously. Considering the nondestructive requirements, the sensors have been positioned reliably and conveniently.

The operating states of turbine-generator sets have direct influence on cavitation. In the sense, the adaptive DAQ and storage is proposed to capture all sound features with less data redundancy. The DAQ period varies with operating states. And different types of data will be stored in association with the operating states and the indexes of cavitation sound.

Based on the real-time evaluation of signal characteristics such as standard deviation, noise level and frequency compositions, the tendency of cavitation intensity with time and different operating states has been traced out. Furthermore, the integrated cavitation intensity will be estimated periodically in terms of the feature of cavitation erosion, the influence of operating states and the water head tendency, which can figure out the degree of cavitation erosion approximately. And the research methodology and pivotal concerns are discussed on the evaluation of the metal loss caused by cavitation.

TrbMAU has been successfully put into on-line operation on 6 turbines in Gezhouba Hydro Power Plant in succession since October 2005. The monitoring and analysis results can describe the cavitation level well. All the achievements are meaningful for carrying out condition-based maintenance of Kaplan turbines.

**Acknowledgments.** Thanks for Professor Zhaohui Li's great supports and instructions.

## References

1. Li, J.W.: Test Technique of Hydraulic Machinery, p. 225. Mechanical Industry Publishing, Beijing (1981)
2. Sirok, B., Kern, I., HoEevar, M., Novak, M.: Monitoring of the Cavitation in the Kaplan Turbine. In: Proceedings of ISIE 1999, Bled, Slovenia, pp. 1224–1228 (1999)
3. Antonio, B.: Real-Time Detection of Cavitations for Hydraulic Turbo-machines. *Journal of Real-Time Imaging* 4, 403–416 (1999)

4. Cudina, M.: Detection of cavitations phenomena in a centrifugal pump using audible sound. *Mechanical Systems and Signal Processing* 17, 1335–1347 (2003)
5. Sheng, P., He, Y.Y., Chu, F.L.: The Online Monitoring and Diagnosis System for Cavitation Erosion of Hydro Turbins. *Water Resources and Hydropower Engineering* (2002)
6. Tian, H., Yu, S.S.: Research on Ultrasonic Detection of Hydraulic Turbine Cavitation. *Nondestructive Testing* (2003)
7. Boja, B.: Vibro-acoustic Method for Diagnosing Cavitation of Hydro Turbines. *EWRI* 18(12) (June 1997)
8. Huang, T.M., Huang, Z.X.: Vibration Features of Hydro Turbine Generators. *Water Resources and Hydropower Engineering* (2003)
9. Shi, H.X., Li, Z.H., Bi, Y.X.: Investigation on Cavitation Characteristics for Hydro Turbines Base on Sound Wave Detection and Its Application. *Water Resources and Hydropower Engineering* (2008)
10. Editorial group of Handbook about Hydro Power Mechanical & Electrical Designing. Handbook about Hydro Power Mechanical & Electrical Designing. China Water and Electric Power Press (1983)
11. Nie, R.S.: Cavitation and Erosion in Hydro Turbines. China Water and Electric Power Press, Beijing (1984)
12. Gordon, J.L.: Hydro turbine Cavitation Erosion. *Journal of Energy Engineering* (118), 194–208 (1992)
13. Tnme, B.A., Li, R.Z., (Interpretation): Optimizing Design of Technical and Economic Parameters for Hydro turbine. Water Resource Press, Beijing (1981)
14. Huang, Z.F.: Research on the Applications of Wavelet Analysis in Cavitation Fault Diagnosis for Centrifugal Pumps. Master Degree Thesis of Jiangshu University (2005)
15. Rose, D.: Noise Principle Underwater. China Ocean Publishing, Beijing (1983)
16. H.И., Ю.У. Cavitation in Hydro turbine. Mechanical industry Publishing, Beijing (1981)
17. Shi, H.X., Li, Z.H., Bi, Y.X.: Investigation on Cavitation Characteristics in Hydro Turbines Based on Frequency Bandwidth Analysis of Wavelet Packet. *Large Electric Machine and Hydraulic Turbine* (2008) (in press)

# Network Capacity Analysis for Survivable WDM Optical Networks

Zhenrong Zhang<sup>1,2</sup>, Zhengbin Li<sup>2</sup>, and Yongqi He<sup>2</sup>

<sup>1</sup> School of Computer, Electronics and Information, Guangxi University, P.R. China

<sup>2</sup> National Laboratory on Local Optic-fiber Communication Networks and Advanced Optical Communication Systems, Department of Electronics, Peking University, Beijing 100871, P.R. China

{zzr76, lizhengbin, heyongqi}@pku.edu.cn

**Abstract.** In this paper, a new architecture of optical networks-the server view based on optical networks is considered. New ILP formulations are proposed to analyze the network capacity with Network survivability under the two types of node servers. Computer simulations are conducted to evaluate the effectiveness of these new formulations. The study has shown that the network can achieve the same throughput under the two types of node servers and the network throughput increases when the maximum allowed variation increases.

**Keywords:** Network survivability, Network Capacity, sever version WDM optical networks.

## 1 Introduction

Wavelength-division multiplexing (WDM) is the favorite multiplexing technology for optical communication networks because it supports a cost-effective method to provide concurrency among multiple transmissions in the wavelength domain [1]. Several communication channels, each carried by a different wavelength, are multiplexed into a single fiber strand at one end and demultiplexed at the other end, thereby enabling multiple simultaneous transmissions. Each communication channel (wavelength) can operate at any electronic processing speed (e.g., OC-192 or OC-768) that is currently available. A WDM optical network would consist of electrically controlled optical nodes. These optical nodes may be optical crossconnect (OXC) (also known as wavelength crossconnects (WXC)) [2], which perform switching of optical wavelength channels. Optical nodes can dynamically configure and reconfigure optical wavelength connections (known as optical lightpaths) between nodes so as to distribute the fiber route resource in the optical layer on the basis of changes in traffic demands. This would then make the best use of network resources.

Characterizing the “capacity” of a WDM optical network has turned out to be a difficult problem owing to the intricacies of communication over fiber links. Integer Linear Programming (ILP) formulations for maximizing network throughput for two types of traffic patterns, uniform and non-uniform was proposed in [3]. Reference [4]

presented algorithms for the design of optimal virtual topologies embedded on wide-area wavelength-routed optical networks. In [5], the authors study the maximum throughput properties of dynamically reconfigurable optical network architectures having wavelength and port constraints. Reference [6] presented a set of scheduling algorithms to maximizing throughput for optical burst switching networks. Capacity of optical networks—the server version WDM optical network [7] was analyzed in reference [8]. In this paper, we proposed ILP formation to analyze the survivable network capacity under two types of node servers, with and without wavelength convertor.

## 2 Server Version Optical Networks

In general, server is a hardware and software device designed to perform a specific function for many users. In information technology, a server is an application or device that performs services for clients as part of a client/server architecture. In networking, server is an object of network that provides a specific service for the network traffic packets. According to this definition, the network resource is categorized into three types—the access server, the node server and the link server, based on the fact that the network resource performs the different function for the traffic packet. Fig. 1 shows the difference of general vision network and server vision network. The difference is that the server vision network treated the network resource as a server system. The server system is consists of some access servers, some node servers and some link servers.

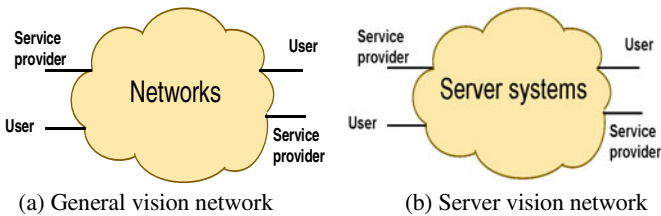


Fig. 1. The difference of general and server vision networks

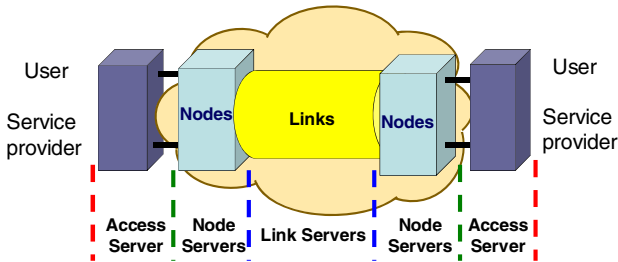


Fig. 2. The server vision network model

The server vision network model is demonstrated in Fig.2. The access server includes IP router servers, Ethernet servers and ATM servers, etc. The IP Router server provides routing and forwarding service for IP packets. The Ethernet server provides forwarding service for Ethernet frames. The ATM server provides switching and routing service for cells. The node server includes optical cross connect (OXC) servers, wavelength convertor (WC) servers and Fiber Delay Lines (FDLs) servers, etc. The OXC server provides service for optical circuits, bursts and packets switching or routing in space domain. The WC server provides service for wavelength channel to resolve contention in wavelength domain. The FDLs server provides service for packet delay some determinate time to resolve contention in time domain. In addition, there are some other servers, such as signaling servers, etc. The link server includes transmitter servers, receiver servers, amplifier servers, and wavelength channel servers. The transmitter server provides data transmission service. The receiver server provides data receiving service. The amplifier server provides signal amplifying service. The wavelength channel server provides channel for signal.

### 3 Capacity Analysis for Survivable WDM Optical Networks

The concept of PWCE, originally introduced by Grover in [9], is attractive because it does not need protection resources to be dynamically configured thereby simplifying network management and reducing processing overheads. For a given network topology with the specified edge (span) capacities, it first assigns some spare capacity to each edge to create an “envelope” of protected working capacities.

In this paper we focus on the total capacity which can be provided the networks. As in [6], we can model a server version WDM optical network as a directed graph  $G(V, E)$  with as the set of nodes  $V$  and  $E$  as the set of directed links. A directed edge transmits data in one direction. Any link in the network may consist of two counter directional links whose capacity may not be equal. We combine the ILP for PWCE design in [10] and the ILP for network capacity in [8], propose the new ILP to analyze the network capacity with PWCE design for survivable WDM optical networks. We use the following notation to present our proposed new ILPs.

$D_{i,j}$ : Demand from node  $i$  to node  $j$  can be provided, in terms of number of lightpaths.

$D_{i,j}^w$ : Demand from node  $i$  to node  $j$  can be provided with wavelength  $w$ , in terms of number of lightpaths.

$P_{i,j}^{s,d,w}$ : The number of Lightpaths from node from node  $s$  to node  $d$  traverses link  $i \rightarrow j$  using wavelength channel  $w$ .

$P_{i,j}^{s,d}$ : The number of Lightpaths from node from node  $s$  to node  $d$  traverses link  $i \rightarrow j$ .

$W$ : The capacity of each link, in terms of number of wavelength channels.

$D_{\max}$  and  $D_{\min}$ :  $D_{\max} = \max\{D_{i,j}\}$  and  $D_{\min} = \min\{D_{i,j}\}$  are the maximum and minimum values of  $D_{i,j}$

$p_{i,j}$  indicates if span  $j$  is an on-cycle span of cycle  $i$  or not;  $p_{i,j}$  is 1 if span  $j$  is an on-cycle span of cycle  $i$ , and 0 otherwise.

$x_{i,r}$  represents the number of backup paths that cycle  $i$  with one unit capacity can provide for the failed span  $r$ ;  $x_{i,r}$  is 1 if the failed span  $r$  is an on-cycle span of cycle  $i$ , 2 if the failed edge  $r$  is a straddling edge of cycle  $i$ , and 0 otherwise.

$w_j$  and  $s_j$  are the number of working and spare units on span  $j$ , respectively.

$\alpha$  is the maximum allowed variation of  $w_j$ .

The capacity of survivable WDM optical networks of node server configured with WCs (wavelength converters) can be formulated as follows:

$$\text{Objective:} \quad \max \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} D_{i,j} \tag{1}$$

$$\text{Constraints:} \quad \sum_{j:(ij \in E)} P_{ij}^{s,d} - \sum_{j:(ji \in E)} P_{ji}^{s,d} = \begin{cases} -D_{s,d}, & i = d \\ D_{s,d}, & i = s \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$\sum_{s \in V} \sum_{d \in V} P_{ij}^{s,d} \leq w_i \quad ij \in E \tag{3}$$

$$s_r = \sum_{i=1}^{|P|} p_{i,r} n_i \quad \forall r \in E \tag{4}$$

$$w_r \leq \sum_{i=1}^{|P|} x_{i,r} n_i \quad \forall r \in E \tag{5}$$

$$w_r + s_r \leq W \quad \forall r \in E \tag{6}$$

$$D_{\max} - D_{\min} \leq \alpha \quad \alpha = 0, 1, 2, \dots \tag{7}$$

Note that  $w_j$  represents the protectable working units on each edge; this is unknown, and will have to be determined by the ILP formulas. The values of  $p_{i,j}$ ,  $x_{i,j}$ ,  $\alpha$ , and the sets of  $P$  and  $E$  can be pre-determined for a given network topology. As in our previously work in [8], these formulations can be easily extended the case of node sever without WCs.

### 4 Numerical Results

To evaluate the effectiveness of these new formulations, we have studied cost239 network and compared the network throughput achieved under two types of node servers (with and without wavelength converter). Here we assume each span to have two fibers transmitting in opposite directions and that there are 16 wavelengths per fiber. The software CPLEX8.0 [11] is used to find the solution for our ILP model. Our investigations show that the network throughput depends on the value of  $\alpha$ .

**Table 1.** Relationship between the network throughput and  $\alpha$ 

$\alpha$	Network capacity	Network Capacity with survivability (with WCs)	Network Capacity with survivability (without WCs)
0	440	330	330
2	546	452	452
4	576	470	470
6	620	532	532
8	670	558	558
10	706	594	594
12	726	640	640
14	780	656	656
16	832	656	656

Table 1 shows the relationship between the network throughput and  $\alpha$ . The results have shown that the network can achieve the same throughput under the two types of node servers and the network throughput increases when the maximum allowed variation of  $D_{ij}$  increases.  $D_{ij}$  means the demand from node  $i$  to node  $j$  can be provided, in terms of number of lightpaths. The networks with survivability can achieve lower capacity compared with the networks without survivability. The ratio of these two models will be around  $[1 - 1/(\bar{d} - 1)]$ , here  $\bar{d}$  is the average node degree of the networks.

## 5 Conclusions

We have presented a new architecture of optical networks—the server version WDM optical network and propose new ILP formulations to analyze the survivable networks throughput under two types of node servers (with and without wavelength converters). To evaluate the effectiveness of these new formulations, we have studied cost239 network and compared the network throughput achieved under two types of node servers. The study has shown that the network can achieve the same throughput under the two types of node servers and the network throughput increases when the maximum allowed variation of  $D_{ij}$  increases.

**Acknowledgments.** This work is partially supported by 973 Program of China (No. 2010CB328202 and 2010CB328203), the Scientific Research Foundation of Guangxi University (XB2100101, X071032).

## References

1. Siva Ram Murthy, C., Gurusamy, M.: WDM Optical Networks: concepts, design, and algorithms. Prentice Hall PTR (2002)
2. Ramaswami, R., Sivarajan, K.N.: Optical Networks: A Practical Perspective. Morgan Kaufmann Publishers (1998)

3. Shiva Kumar, M., Sreenivasa Kumar, P.: Static lightpath establishment in WDM networks- New ILP formulations and heuristic algorithms. *Computer Communications*, 109–114 (2002)
4. Banerjee, D., Mukherjee, B.: Wavelength-Routed Optical Networks: Linear Formulation, Resource Budgeting Tradeoffs, and a Reconfiguration Study. *IEEE/ACM Transactions on Networking*, 598–607 (October 2000)
5. Brzezinski, A., Modiano, E.: Achieving 100% Throughput in Reconfigurable Optical Networks. *IEEE/ACM Transactions on Networking*, 970–983 (August 2008)
6. Li, J., Qiao, C., Xu, J., Xu, D.: Maximizing Throughput for Optical Burst Switching Networks. *IEEE/ACM Transactions on Networking*, 1163–1175 (October 2007)
7. Zhang, Z., Li, Z.: Server Vision: A New Perspective for Next Generation Optical Networks. In: *Proceeding of SPIE-Apoc. 2008, Hangzhou, China, October 26-31 (2008)*
8. Zhang, Z., Li, Z., Xu, A.: Network Capacity Analysis of Server Version WDM Optical Networks. *China Communications* 7, 89–93 (2009)
9. Grover, W.D.: The protected working capacity envelope concept: an alternate paradigm for automated service provisioning. *IEEE Communication Magazine* 42, 62–69 (2004)
10. Zhang, Z.R., Zhong, W.D., Bose, S.K.: Dynamically survivable WDM networks design with p-cycle-based PWCE. *IEEE Communications Letters* 9, 756–758 (2005)
11. <http://www.ampl.com/DOWNLOADS/cplex81.html>



# Study on Traffic Information Dissemination Scheme in Transit of Highway Network

Jun-jun Tang

8 Xitucheng Rd, Beijing, China  
jj.tang@rioh.cn, tangjjming@126.com

**Abstract.** Traffic information's effective dissemination can improve traffic operation condition of highway network, and expand its traffic capacity. From the angle of highway network, this paper studies on various ways to disseminate guidance information to traffic flow by roadside variable message board, highway advisory radio and intelligent terminal, and puts forward the setting principle and releasing content of them, which all will increase the timeliness, completeness and effectiveness of traffic information released.

**Keywords:** Highway network, Traffic information, Dissemination way, Variable message board.

## 1 Introduction

America, Japan and other developed countries have developed Intelligent Transportation System (ITS) early, where the scale and speed of development are astonishing. Information dissemination technology is one of important components of ITS.

The United States have already established traffic information dissemination system based on GIS, which can provide all-dimensional traffic information service for traffic managers and users, such as traffic congestion, unexpected events, video surveillance images and other contents, by means of management center, variable message board, radio and television, the Internet, mobile terminal, etc. In addition, America has also set up a comprehensive 511 traffic management and information platform, which can inquire real-time traffic information (Figure 1).

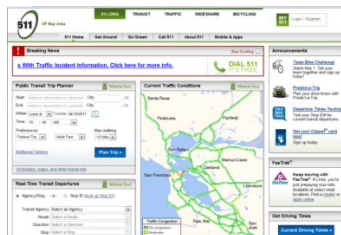


Fig. 1. 511 traffic management and information platform

Japan has formed an advanced Vehicle Information and Communication System (VICS), and its information dissemination ways are mainly radio wave beacons, light beacons and FM multiplex broadcasts, which can provide a guarantee for traffic safety management of highway network.

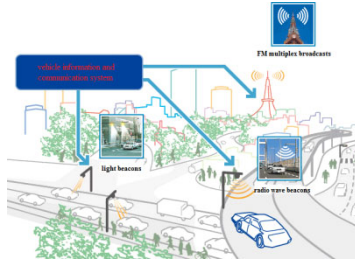


Fig. 2. Information dissemination ways of VICS

According to actual road construction and traffic situation in China, this paper puts forward the following using scenes of various traffic information dissemination ways, and forms a comprehensive scheme in transit of highway network. At last, an illustrative example is given.

## 2 Major Traffic Information Dissemination Ways

Traffic information dissemination is mainly to communicate road condition, traffic situation, vehicle running and traffic service information through variable message boards, variable speed limit signs, highway advisory radio, vehicle or handheld intelligent terminal, which will help travelers to reach the destination in best travel ways and routes. Basically, some major ways are as follows.

### Normal Traffic Information Dissemination

- Variable message board

Variable message boards are still the most widely used to release traffic information on expressways and urban express trunk roads, which can be installed at exits, toll stations and main lines of expressways, intersections of expressways and ordinary roads, and are applicable to release accidents, incidents, meteorological, traffic guidance, control and other information. Their specific applications are varied and as shown from Figure 3 to Figure 7.



Fig. 3. Road weather information



Fig. 4. Road congestion information



Fig. 5. Road closed information



Fig. 6. Travel time information



Fig. 7. Traffic incidents information

● Variable message sign

By comparison with variable message boards, variable message signs are limited in contents released, but have reliable real-time, which are shown in Figure 8 to Figure 9.



Fig. 8. Lane control information



Fig. 9. Variable speed limit information

**Emergent Traffic Information Dissemination**

● Mobile variable message board

The function of mobile variable message board is similar to roadside small variable message board. When the expressway appears adverse weather, road-works, emergencies which are difficult to be excluded in a short time, some traffic management and control information must be communicated timely to drivers to ensure traffic safety. While there is no fixed variable message board in the vicinity, mobile variable message board can be used right here and now. It is small, light and flexible with a variety of display and power supply mode, and can be installed on the ground or a vehicle anywhere and anytime.



**Fig. 10.** Mobile variable message board

- Traffic weather automatic trigger system

Traffic weather automatic trigger system includes a visibility sensor, a trigger device, flashing lights and a variable message board. When monitoring the visibility of a road section using a visibility sensor, the system can set a threshold for the visibility. And if the visibility drops below that threshold, the system will trigger and close the relay. Then this relay can be used to turn on flashing lights and a variable message board. This provides real-time weather condition information to drivers.



**Fig. 11.** Traffic weather automatic trigger system

### Auxiliary Traffic Information Dissemination

- Highway advisory radio system

Highway advisory radio system includes stations, flashing lights and a static message sign. The system alerts drivers to potential dangers with a combination of flashing lights and a static message sign instructing drivers to tune their radios to a particular frequency for important announcements, such as traffic situation, weather events, emergencies and evacuation information on the road, in order to create a safer and smoother commute and maintain a safe driving environment. In a word, the system can assist in managing traffic during long-term construction projects, and keep traffic flowing during special events. It also can simply be used for frequently congested traffic areas.



**Fig. 12.** Highway advisory radio system

- Short message service (SMS)

SMS is a supplementary way for release temporary traffic information. Drivers can register before departure, and then they will receive timely traffic information on roads.

**New Traffic Information Dissemination**

- Intelligent terminal

Intelligent terminal is an acceptable terminal display and query device. Combining of GPS and GIS technology, it communicates with wireless communication devices installed on expressways, and displays the current vehicle position, traffic situation ahead, and reasonable travel routes with graphics and text. Although vehicle or handheld intelligent terminal can timely reflect road condition, they are rarely used because they have some constraints based on the price and integrated information platform which is not established effectively yet.



Fig. 13. Mobile vehicle road condition television in China

### 3 Traffic Information Dissemination Scheme in Transit

As shown in Figure 14, there is a road network structure map, in which black lines represent urban express trunk roads, grey lines represent a ring road and expressways. In response to traffic information dissemination ways above, based on the vehicle's location in the road network, a complete set of traffic information dissemination scheme is as follows.



Fig. 14. Road network structure map

**Driving from Urban Express Trunk Roads to the Ring Road**

It is a suggestion that variable message boards can be installed on urban express trunk roads, 1-2 kilometers ahead of the entrance to the ring road, which are used to release traffic situation of urban express trunk roads and the ring road, as shown in Figure 15.



Fig. 15. The scheme of driving from urban express trunk roads to the ring road

### Driving from Expressways to the Ring Road

It is a suggestion that variable message boards can be installed on expressways, 1-2 kilometers ahead of the entrance to the ring road, which are used to release traffic situation of the ring road and other expressways, as shown in Figure 16.

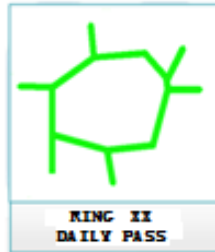


Fig. 16. The scheme of driving from expressways to the ring road

### Driving from the Ring Way to Expressways or Urban Express Trunk Roads

It is a suggestion that variable message boards can be installed 0.6-0.8 kilometers ahead of the exit to the ring road, which are used to release traffic situation of expressways and urban express trunk roads connected to the ring road, as shown in Figure 17.



Fig. 17. The scheme of driving from the ring way to expressways or urban express trunk roads

### **Driving on the Ring Way**

It is a suggestion that variable message boards can be installed 1-2 kilometers of the upper reaches of transportation junctions of the ring road, which are used to release various incidents, guidance information and travel time ahead, as shown in Figure 6.

### **Driving on Expressways**

It is a suggestion that variable message boards can be installed ahead of the entrance to toll stations, which are used to release traffic situation of expressways (Figure 7), and special traffic information query terminal can be located in filling Stations, gas stations and important entrances, applying together with highway advisory radio system, SMS, mobile variable message board, traffic weather automatic trigger system, and other traffic information dissemination ways.

## **References**

1. Chen, B.: Intelligent Transport System Development and Current Status in America. *Road Traffic & Safty*, 36–38 (2003)
2. Chen, H.: The Traffic Management System for Osaka—Kobe Expressway in Japan. *Shanghai Coustruction Science and Technology*, 12–15 (1996)
3. Shi, L.-J., Zhu, J., Chen, X.-H.: Study on the Traffic Guidance Scheme of Road Network in Pudong New Zone. *East China Highway*, 3–6 (2010)
4. Vaisala web site. Highway Advisory Radio, <http://www.vaisala.com/en/products/highwayadvisoryradio/Pages/default.aspx>

# Analysis and Emulation of a Layered Aggregation Method Based on Fair

Youjun Chen, Yunbing Zeng, and Hongying He

College of Mathematics and Information, China West Normal University,  
Nanchong, Sichuan, China

**Abstract.** In order to avoid affection of aggregation frequency and get delay guarantee, a layer-by-layer aggregation method based on fair is proposed. Gives complete theoretical analysis for this method. The experiment result based on NS2 shows that the method can provide the better delay performance.

**Keywords:** layer-by-layer aggregation, delay performance, fair aggregation.

## 1 Introduction

For the non-bitwise operations packet-switched networks, the main problem is package accumulation while aggregation. At the both ends of layered aggregation are network nodes that have equivalent level, aggregation occurred in each of up-link, and solution aggregation occurred in each of down-link. As the more of layers and times of aggregation, the delay times obviously increased, reference [1-2] pointed out hierarchy aggregation can achieve better delay performance on fair, but didn't give an aggregation algorithm. This paper proposes an aggregation and solution aggregation algorithm on fair, then, suggest a double-layer preemptive aggregation and fair solution aggregation method.

## 2 Algorithms of Aggregation and Solution Aggregation Based on Fair

The new algorithm adopted VC scheduling algorithm [1] based on router packet scheduling algorithm. The VC algorithm is a statistical time division service model based on TDM, that can effectively meet a certain service rate guarantee, delay guarantee and flow isolation monitor.

### Double-Layer Preemptive Aggregation

Aggregation points two layers, the upper layer are macro flows after aggregation that run the fair queue algorithm [1-5]. The lower layer is composed of constitute streams of each macro flow, they connected respectively with the corresponding macro flow queue, and also separately run the fair queue algorithm.



The reason of using two layers are to guarantee the fairness of aggregation and make the output flow after aggregation to meet the constraints of services rate of macro flow. Each macro flow queue at most keep one packet that maintains the virtual times of macro flows, but this packet come from the stream queue of the bottom layer, and the queue of the bottom layer maintains it's own virtual times, so, every layer has it's own different virtual times.

The packets in macro flow queue must maintain two virtual times, the own virtual time of macro flow is the base of scheduling between macro flows, but the virtual times from the flow queue of bottom layer is the base of replacement. If there is a packet come from the bottom layer queue and its virtual time is smaller than the packet that is in a waiting state of upper layer macro flow, then, exchange the two packets, and virtual times of macro flow queue as a corresponding adjustment.

This replacement is preemptive, so said the aggregation is a double layer preemptive aggregation. The replacement helps to maintain the fairness of aggregation and reduce the additional aggregation delay brought by unfairness.

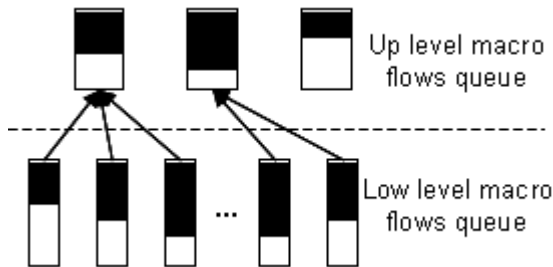
**Solution Aggregation**

Because the packets of different macro flows may reach the same output link, so, in order to achieve a fair solution aggregation, the system need to run the queue for each macro flow. In addition, because the packet from the macro flow can direct enter into a single flow queue, so the waiting time for packet is the time that in the solution aggregation queue, it is consistent with VC delay boundaries.

On the other hand, some packets transmitted in the domain may not satisfy the rules of the constraints when there are some changes of traffic characteristics, but these packets have a certain delay boundaries in each node, so, the packets can only be to arrive early, then, deduct the ahead of arrival times, the packets satisfy the rules of constraints.

**Structures of Aggregator and Solution Aggregator**

In summary, we can give the network structure of a layer-by-layer aggregation based on fair.



**Fig. 1.** Schematic of aggregator

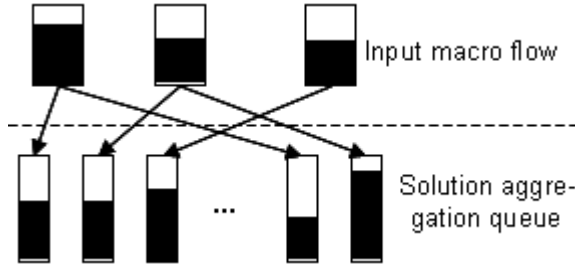


Fig. 2. Schematic of solution aggregator

The figure 1 is the structure of double layer preemptive aggregator, the figure 2 is the structure of solution aggregator based on fair.

### 3 Performance Analysis

First, we give the relevant definition of some basic terminology: Let  $f$  be micro flow,  $p_f$  be the packets of  $f$ , and  $p_f^i$  be the packet of number  $i$ ,  $l$  be the length of packet,  $F$  be aggregation flows,  $\max(l_f)$  and  $\max(l_F)$  be the max length of packets of  $f$  and  $F$  in corresponding. Flows subject to  $(\sigma, \rho)$  model,  $\sigma$  be the burst size,  $\rho$  be the flows rate, and  $\rho_f$  be the rate of flow  $f$ ,  $\sigma_f$  be the burst size of flow.  $R$  (It can also be  $r$ ) be the service rate of scheduling algorithms, such as  $R_f$  be service rate of flow  $f$ , and also  $R_f \geq \rho_f$ . Let  $h$  be the nodes on the path from entrance to exit,  $\max(l^n)$  be max length of all flows in node  $n$ , let  $C_n$  be the output link bandwidth of node  $n$ . Let  $d$  be the delay time,  $D$  be the worst delay time,  $D_f$  be the worst delay time of flow  $f$ , and  $D(p)$  be the delay boundaries.

By using the PGS flow scheduling method based on the flow rate, the worst delay time of flow  $f$  which satisfies the traffic flow constraint  $(\sigma_f, \rho_f)$  can be expressed as:

$$D_f = \frac{(\sigma_f + h * \max(l_f))}{\rho_f} + \sum_{n=1}^h \frac{\max(l^n)}{C_n}$$

Similarly, by using aggregation scheduling methods, the worst delay time can be expressed as:

$$D_f(F) \leq (h-2) \left( \frac{\max(l_F)}{\rho_F} + \frac{\sum_{n=2}^{h-1} \max(l^n)}{C_n} \right) + d_f^{in} + d_f^{out}$$

In this expression,  $d_f^{in}$  is the additional delay time of flow  $f$  when aggregation,  $d_f^{out}$  is the additional delay time of flow  $f$  when solution aggregation,  $h-2$  is the queue delay time at entrance nodes and exit nodes.

Assuming that each incoming flow is a rule flow and which satisfies  $A(p_i^j) \geq D(p_i^{j-1})$ , among this expression, let  $p_i^j$  be the  $j$ th packet of flow  $i$ ,  $A(\cdot)$  be the arrival time,  $D(\cdot)$  be the delay boundary time. Therefore, the time of any packet  $p_i^j$  which length is  $l_i^j$  that last time (not to be replaced) enter into the macro flow queue satisfy:

- (1) Some packets are being sent while  $p_i^j$  enter into flow  $F$

$$A_F(p_i^j) \leq A(p_i^j) + \frac{l_i^j}{r_i} - \frac{l_i^j}{r_F} + \frac{\max(l_F)}{C}$$

- (2) Flow  $F$  is in an idle state

$$A_F(p_i^j) \leq A(p_i^j) + \frac{l_i^j}{r_i} - \frac{l_i^j}{rF}$$

This is because flow  $F$  can be viewed as an output link that satisfies the delay guarantee of VC.

After  $p_i^j$  entered flow  $F$ , the transmission time is:

- (1) Some packets are being sent while  $p_i^j$  enter into flow  $F$

$$E(p_i^j) \leq A_F(p_i^j) + \frac{l_i^j}{r_F} \leq A(p_i^j) + \frac{l_i^j}{r_i} + \frac{\max(l_F)}{C}$$

- (2) Flow  $F$  is in an idle state

$$A_F(p_i^j) + \frac{l_i^j}{r_F} + \frac{\max(l_m)}{C} \leq A(p_i^j) + \frac{l_i^j}{r_i} + \frac{\max(l_m)}{C}$$

Because  $\max(l_F) \leq \max(l_m)$ ,  $l_m$  is the max length of entire node (output link), so, there will be:

$$E(p_i^j) \leq A(p_i^j) + \frac{l_i^j}{r_i} + \frac{\max(l_m)}{C}$$

That is to say the delay time of packet  $p_i^j$  is:

$$d_{in} \leq \frac{l_i^j}{r_i} + \frac{\max(l_m)}{C}$$

Similarly, we can get that,  $d_{out} \leq \frac{\max(l_f)}{r_f} + \frac{\max(l_m)}{C}$  so, the worst delay time of flow  $f$  aggregate to macro flow  $F$  is:

$$D_f(F) \leq (h-2) \frac{\max(l_F)}{\rho_F} + 2 \frac{\max(l_f)}{\rho_f} + \frac{\sum_{n=1}^h \max(l^n)}{C_n}$$

By comparing  $D_f$  and  $D_f(F)$ , we can get the Delay difference:

$$\Delta d_f = D_f(F) - D_f = (h-2) \left( \frac{\max(l_f)}{\rho_f} - \frac{\max(l_F)}{\rho_F} \right)$$

As a result, we can get the condition of improving the worst delay performance:

$$\frac{\max(l_f)}{\rho_f} > \frac{\max(l_F)}{\rho_F}$$

Therefore, the key of improving the worst-case delay performance is the ratio of maximum packet length and the service rate. Commonly, there have  $\rho_F > \rho_f$ , and also, the max length of packets is limited (E.g., by the Ethernet maximum packets length limit), so, compared in terms of per-flow scheduling, flow aggregation can often improve the delay performance at worst case.

From the above results can be seen, the performance of layer by layer aggregation algorithm based on fair at the aggregation node will not be worse than the per-flow.

## 4 Simulation Results Based on Ns2

Layer by layer aggregation and edge aggregation are all based on the path, they have many similar on delay performance [5-8]. The simulation experiments mainly analysis two aspects: one is the delay performance comparison of layer by layer aggregation and edge aggregation, the other is performance influence of aggregation methods based on fair to layer by layer aggregation.

By using free network simulation software NS2 [9], utilizing a four-level network model to the simulation experiment. Topology shown in Figure 3, the different colors show different levels of nodes in the network.

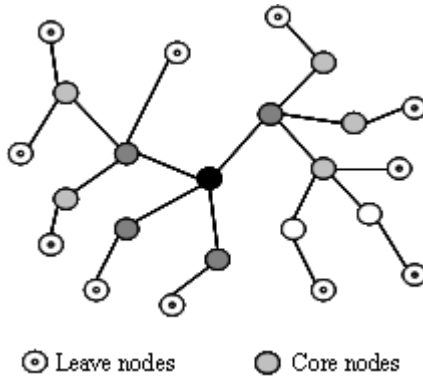


Fig. 3. Network topology

In this experiment, we set a static CBR traffic that can be sent at a steady rate in the entire experiment period between any two edge nodes, then, we can compare the delay performance of three aggregation methods. Figure 4 shows the delay performances of a single flow by using three different scheduling methods. Three columns in the figure 4 represent the delay times (max delay time, minimum delay time and average delay time) of edge aggregation, layer by layer aggregation based on fair and ordinary (unfair) aggregation.

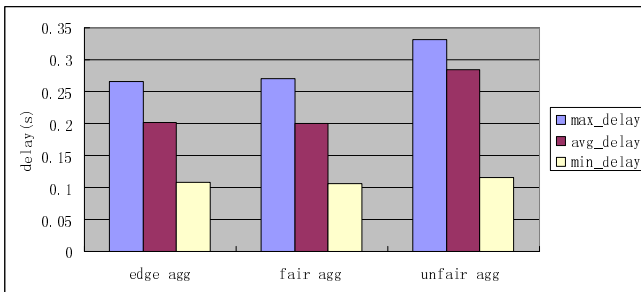


Fig. 4. Delay performances of different aggregation methods

Figure 4 shows the edge aggregation method has the best delay performance, but for the two methods by using layer by layer aggregation method, the method based on fair has a better maximum delay performance and average delay performance, this is because fair method can avoid influences of traffics of flows each other. Furthermore, figure 4 also shows the edge aggregation method and layer by layer aggregation based on fair can get a very close delay performance, so, layer by layer aggregation based on fair can achieve a better delay performance.

## 5 Conclusions

Although the mathematical model of layer by layer aggregation method seems to be more complex, but in fact, it can be simply achieve. Which method used in the network determined by the ISP and it based on status of the network, but the best way is using layer by layer aggregation method if the scale of network is larger, although this method expend some delay performance, but it can provide a better scalability.

**Acknowledgments.** A Project Supported by Scientific Research Fund of Sichuan Education Department (10ZB018).

Youjun Chen is with College of Mathematics and Information, China West Normal University, Nanchong, Sichuan 637009, P. R. China. (Phone: +86-18990777886; Fax: +86-8172568505; E-mail: chenzyw@126.com)

Yunbing Zeng is with College of Mathematics and Information, China West Normal University, Nanchong, Sichuan 637009, P. R. China. (E-mail: ncyzb@sohu.com)

Hongying He is with College of Mathematics and Information, China West Normal University, Nanchong, Sichuan 637009, P. R. China. (E-mail: heyyang@126.com)

## References

1. Cobb, J.A.: Preserving Quality of Service Guarantees in Spite of Flow Aggregation. *IEEE / ACM Transactions on Networking* 10(1), 43–53 (2002)
2. Schmitt, J., Karsten, M., Steinmetz, R.: On the aggregation of guaranteed service flows. *Computer Communications* 24(1), 2–18 (2001)
3. Ergen, S.C.: Energy efficient routing with delay guarantee for sensor networks. *Wireless Networks Archive* 13(5), 679–690 (2007)
4. Xie, G.G., Lam, S.S.: Delay Guarantee of Virtual Clock Server. *IEEE / ACM Transaction on Networking* 3(6), 683–689 (1995)
5. Yang, M., Qian, H.: Together with the path-based scheduling to improve the flow delay performance. *Microelectronics and Computer* (2), 32–36, 54 (2002)
6. Guerin, R., Peris, V.: Quality-of-service in packet network: basic mechanisms and directions. *Computer Networks* (31), 169–189 (1999)
7. Ji, Y., Wang, H.: *Optical burst switching networks*. Beijing University of Posts and Telecommunications Press, Beijing (2005)
8. Chen, M.: *Modern network system theory and technology*. Mechanical Industry Press, Beijing (2007)
9. Wang, Y., Jiao, Y., Meng, T., Lin, J.: Ns based on free software, wireless network simulation. *Radio Engineering* 34(1), 18–20 (2004)

# MIP: A New Routing Metric for Wireless Mesh Networks

Sufen Zhao

The Dept. of Computer Science, HuaZhong Normal University,  
Wuhan, Hubei, China 430079  
Sufen\_Zhao@163.com

**Abstract.** A new simple and effective routing metric called MIP (Metric of Interference and Propagation Delay) for multi-channel multi-radio wireless mesh networks (WMNs) has been proposed. The new metric combines loss rate, channel capacity, inter-flow interference and intra-flow interference together to find effective routing paths for source-destination pairs in WMNs. MIP is different from any existing routing metrics for wireless networks. It redefines Interference Ratio and tries to obtain balance between channel diversity and path length. Large numbers of experimental rounds prove that the new metric improves the performance of WMN a lot than existing algorithms.

**Keywords:** Wireless Mesh Network, routing, MIP, multi-channel multi-radio, interference ratio.

## 1 Introduction

Wireless mesh networks (WMNs) have gained much attention in recent years. It is a promising technology for numerous applications for it can provide high-speed last-mile connectivity, supporting community networks, and so forth. The persistent driving force comes from the techniques itself including extended coverage, robustness, self-configuration, easy maintenance and low cost[1][2][3][4].

However, due to the MAC contention, wireless nodes in the same interference range can not be active simultaneously. As a result, the throughput of WMNs decreases greatly. IEEE 802.11 now can provide multiple channels and multiple NICs (Network Interface Cards) to increase the network capacity with standard hardware. This can really alleviate the bandwidth contention problem. Generally speaking, the number of interfaces at a host is 1-3, while the number of channels available is up to 12 (IEEE 802.11a/b/g) in the network. See an example in Fig. 1. In this figure, each node is equipped with 2 wireless NICs, the network as a whole uses 5 distinct channels. In such multi-channel multi-radio wireless mesh network, two wireless links can transmit data simultaneously if they use different channels even though they are in each other's interference range. Therefore, the bandwidth utilization and the overall throughput of the network are improved.

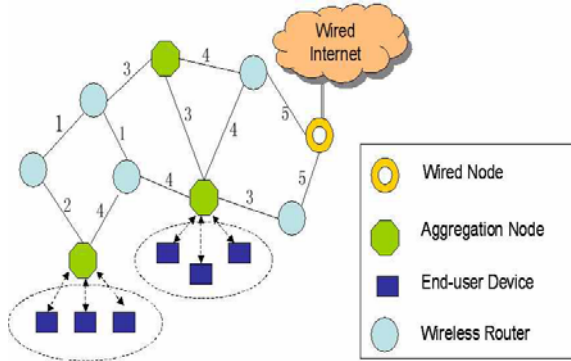


Fig. 1. A WMN example

However, in such multi-channel multi-radio WMNs, how to allocate channel to each interface and find an optimal routing path for each source-destination pair are challenging problems. These two problems are believed to be inter-dependent problems [5][6]. In this essay, we only focus on the latter problem. In other words, we try to find optimal routing paths for WMNs which channels and interfaces have already been assigned.

As far as we know that some high-throughput routing metrics for such WMNs have been proposed, such as ETX [7](Expected Transmission Count), WCETT [8](Weighted Cumulative ETT) , MIC [9](Metric of Interference and Channel-Switching) and iAWARE [10] (In terference Aware Routing Metric) metrics, etc. These previous metrics take the various characteristics of WMNs for account and could obtain better performance than Shortest Path Tree algorithm. But these metrics also have shortcomings, more effective routing metric is highly desirable.

In this paper, we propose a new metric called MIP (Metric of Interference and Propagation Delay) for such WMNs which channels and interfaces have already been allocated. MIP combines the path length, link capacity, link loss rate and the interference together to find a high throughput path for each source destination pair. Large quantities of simulations show that it has better performance than SPT and WCETT.

## 2 New Routing Metric Related Works

Before we introduce MIP, we review several existing routing metrics briefly, such as ETX and ETT.

De Conto proposed a routing metric which called ETX[7]. ETX measures the expected number of transmissions, including retransmissions needed to send unicast packet across a link. Let  $p_l$  denotes the failure probability of transmission for a link  $l$ , then the expected number of transmissions required to successfully deliver a packet over link  $l$  is denoted by [8]:



$$ETX_l = \frac{1}{1 - p_l} \quad (1)$$

Using ETX as a routing metric, one can find the paths which have relatively higher reliability and shorter length. However, in WMNs, many other factors should be taken into account.

Based on ETX, Richard Draves proposed a new concept which called ETT [8]. Assume  $S$  is the packet size,  $B_l$  is the bandwidth of the link  $l$ . the expected transmission time(ETT) for a single packet is denoted as:

$$ETT_l = ETX_l * \frac{S}{B_l} \quad (2)$$

ETT combines link loss rate and channel capacity together to measure the quality of the link. Using ETT as a routing metric can assure finding more reliable and wider paths.

However, when there is no interference in the network, ETT metric can capture the quality of links quite well. But when there are interference in the network, that is not the case. There are two kinds of interference in WMNs: inter-flow interference and intra-flow interference. The former means interference between different flows while the latter means interference caused by one flow travelling along different links. WCETT, MIC, and iAWARE take one or both interferences into account for finding routing path, but they are inadequate.

### 3 New Routing Metric: Mip

In this section, we will propose a new routing metric which called MIP to capture the characteristics of WMNs.

Let  $I(l)$  denotes the interference link set for a link  $l$ . For 802.11 MAC protocol, the interference links for the link  $l$  is the links which are within two hops away from the link and meanwhile using the same transmission channel as link  $l$ . Let  $R_l$  be the average total flow rate on link  $l$ , that is, the average link load. Then we define a new concept  $\tau_l$  as:

$$\tau_l = \frac{R_l}{B_l} \quad (3)$$

One can see that the value of  $\tau_l$  gives the fraction of time for the total flow occupies the bandwidth of the link  $l$ . Its value is between 0 and 1, and it reflects the average link load status for link  $l$ .

If a lot of links are in the interference range of link  $l$ , the interference of  $l$  is likely to be heavy. However, if the interference links have small packets to transmit they will not affect link  $l$  too much. Therefore, the interference severity of a link is more related to the interference links' load status than to the number of interference nodes around the link. Hence, we define the Interference Ratio (IR) for link  $l$  as:

$$IR_l = \frac{\sum_{l' \in I(l)} \tau_{l'}}{1 + \sum_{l' \in I(l)} \tau_{l'}} \tag{4}$$

1 in the denominator means the assumptions that link  $l$  will transmit data at full speed. IR means the interference severity of link  $l$ , and it shows the time slices for the interference links while link  $l$  is willing to transmit data at full speed. The larger the value of IR, the more serious the interference of the link. The definition of IR is different from any other existing routing metrics. Its' definition is more scientific and could be calculated more easily than in iAWARE.

When there is interference around link  $l$ , time for transmitting a single packet successfully will increase. How can we calculate the new expected transmission time for a single packet?

For IR denotes the time slices for interference links,  $1-IR$  will be the time slices for link  $l$  transmitting.  $1-IR$  equals  $\frac{1}{1 + \sum_{l' \in I(l)} \tau_{l'}}$ . Thus, the expected time for successfully transmitting a single packet now is revised as:

$$t_l = ETT_l \div \frac{1}{1 + \sum_{l' \in I(l)} \tau_{l'}}$$

The deformation of the above equation is as :

$$t_l = ETT_l \times (1 + \sum_{l' \in I(l)} \tau_{l'}) \tag{5}$$

The formula (5) gives the expected transmission time of a single packet while we take the inter-flow interference into account.

However, except for inter-flow interference, we should also consider intra-flow interference. For example, assuming there are two paths  $l_1$  and  $l_2$  from source  $S$  to destination  $D$ . The sum of  $t_l$  weight for path  $l_1$  is a bit larger than  $l_2$ , but path  $l_2$  chooses more diverse channels than  $l_1$ . In this case, we believe that  $l_2$  is better than  $l_1$  because intra-flow interference of  $l_2$  is expected to be lighter than  $l_1$ , and transmission delay on  $l_2$  will be less than on  $l_1$ . In a word, a path that is made up of hops on different channels is better than a path where all the hops are on the same channel.

Inspired by WCETT, we assume that one  $n$  hops path  $p$  chooses  $m$  different channels. Let  $ch(l)$  denotes the channel selected by link  $l$ . Links which use the same channel can not transmit data simultaneously while the others can. So we can add up the value of the transmission time of the links (here it is the  $t_l$ ) which use the same channel. Then, we will get  $m$  different sum values, each of which is the total transmission time on each channel. Among the  $m$  different channels, there will be a

bottleneck channel which has the largest sum value. The total transmission time on the bottleneck channel is an approximation of the total transmission time on path  $p$ . That is:

$$t_p = \max_{1 \leq j \leq m} \sum_{l \in p, ch(l)=j} t_l \tag{6}$$

The value of  $t_p$  has the same meaning with the max part of WCETT. But like WCETT, the above metric will not always increase as more hops are added to the path. This is a problem. From [8], we know that a good routing metric should satisfy the condition that if we add a hop to the existing path, the cost of the path should increase. In order to solve the problem, we add an additional part which called PT to each hop. PT means the propagation delay. Thus, for a  $n$  hop path  $p$ , we define the new path metric as:

$$MIP(p) = \max_{1 \leq j \leq m} \sum_{l \in p, ch(l)=j} t_l + \sum_{i=1}^n PT_i \tag{7}$$

This metric is MIP. One can see that MIP can estimate the total transmission time for the packets while using  $p$  as the routing path. The two parts of formula (7) can be treated as the balance between channel-diversity and path length. The newly added part could assure using Dijkstra’s algorithm to find the optimal routing paths.

The newly proposed routing metric MIP is different from any existing routing metrics for WMNs. Compared with WCETT, it captures the inter-flow interference because it favors a path that occupies less channel time at its neighboring nodes and thus is believed to obtain more throughput in real WMNs.

### 4 Analysis for the Routing Metric

From [11], we know that the correctness of Dijkstra’s algorithm to compute S-lightest path should be under the assumption of isotonicity of the algebra. Let the algebra include the set  $S$ , the operation  $\oplus$  and the order relation  $\preceq$ . The isotonicity is the most important property of the algebra, that is  $P5$  property [12]:

$\oplus$  is isotone for  $\preceq$ :  $a \preceq b$  implies both  $a \oplus c \preceq b \oplus c$  and  $c \oplus a \preceq c \oplus b$  for all  $a, b, c$  and  $S$ .

Isotonicity is the necessary and sufficient condition for using Dijkstra’s algorithm or Bellman-Ford algorithms to compute S-lightest path. Unfortunately, all the above metrics for multi-channel multi- radio WMNs include WCETT, iAWARE, MIC and MIP metrics don’t satisfy this constraint. As far as we know that, in the absence of isotonicity, using any greedy algorithm can not assure finding the lightest path at any case, except for using enumeration methods. However, enumeration method has the unbelievable high complexity.

However, even though MIP is not isotonic, it can still assure that one can always find a loop free routing path if not by hop-by-hop manner. From large times of simulation rounds, we find that the algorithm could find the best routing paths in most cases and has good overall performs.

## 5 Simulation Results

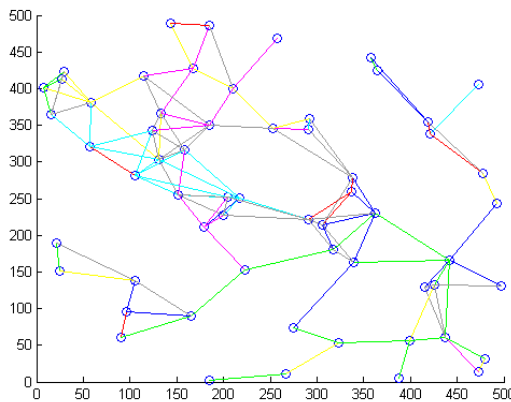
**Table 1.** The parameters for simulation

Parameter	Value
Channel number	7
Nic Number	3
Node Number	40,50,60,70,80
Packet Loss Ratio	0-0.8
Bandwidth	5-20M
Packet Size	1024bit
Propogation Delay	0.0001s, 0.00015s
$\beta$ (WCETT)	0.5

We perform simulation using MATLAB tool. The simulation is partitioned to several steps.

Firstly, we use simulator to form a wireless mesh network structure. The Zigbee alliance use IEEE 802.15.4 as a basis to propose the specifications for network layer, application layer and security of LR-WPANs. The proposal is based on mesh tree approach, that is firstly construct An Adaptive Robust Tree(ART) for WMN, then form a Meshed Adaptive Robust Tree(MART) under some rules. We utilize this approach to form the wireless mesh backbone network.

The second step is to allocate channels to each virtual wireless links. we implement neighbor-partition method on he generated mesh network. The number of channels and interfaces will have great impact on the performance of WMNs described in [3]. Fig.2 is an example of network model whose channels has already been allocated for the network. The number of channels is 7 and the NIC of each node is 3. Different color means different channels.



**Fig. 2.** A network model example with 60 nodes (different color means different channels)

The third and the most important step is to implement our routing algorithm on the network which channels have already been allocated. We also generate traffic for each virtual wireless links and compute the shortest path using Dijkstra’s algorithm. Table 1 gives the parameters we used. However, we also implement shortest path tree (SPT) and WCETT routing algorithms on the WMNs we generated and compare the average path metric for them.

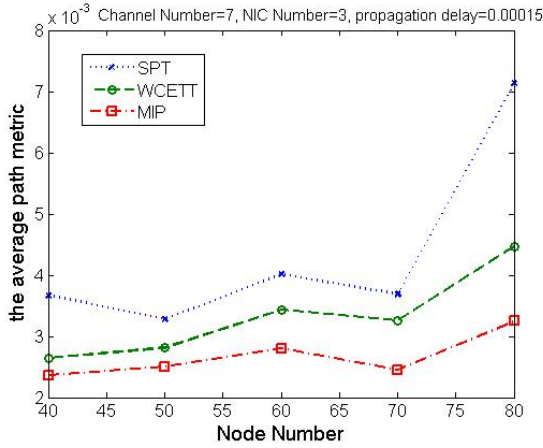


Fig. 3. The average path metric for different routing algorithms with propagation delay=0.00015s

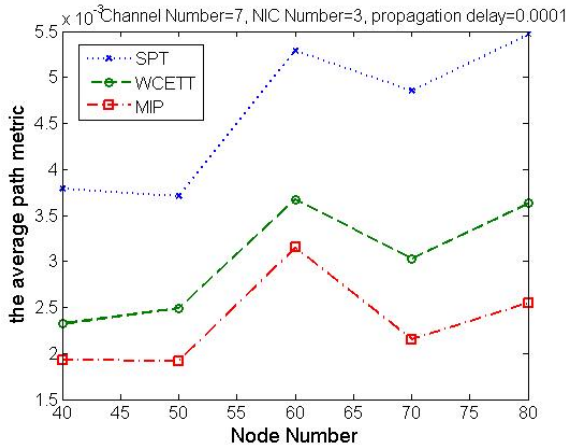


Fig. 4. The average path metric for different routing algorithms which propagation delay=0.0001s

The simulation results is presented as in Fig. 3 and fig.4. The whole simulation run 15 times for each configuration. From Fig. 3 and Fig. 4, one observes that in case of propaga-on delay equal to 0.0001s and 0.00015s, our algorithm can obtain lower average path metric than SPT and WCETT.

## 6 Conclusion and Future Work

In this paper, we introduce a new routing metric MIP for multi-channel multi-interface wireless mesh networks. Simulation results show that MIP can obtain better performance than SPT and WCETT. In the future, we will implement our algorithm on NS or OPNET and obtain the throughput metric to prove the significance of our metric.

## References

1. Akyildiz, I.F., Wang, X.: A survey on wireless mesh networks. *IEEE Communication Magazine* 43(9), S23–S30 (2005)
2. Kyasanur, P., So, J., Chereddi, C., Vaidya, N.H.: Multichannel mesh networks: challenges and protocols. *IEEE Wireless Communications* 13(2), 30–36 (2006)
3. Subramanian, A.P., Buddhikot, M.M., Miller, S.: 2nd IEEE Workshop on Wireless Mesh Networks, WiMesh 2006, September 25-28, pp. 55–63 (2006)
4. Lee, M.J., Zheng, J., Ko, Y.-B., Shrestha, D.M.: Emerging standards for Wireless mesh technology. *IEEE Wireless Communications* 13(2), 56–63 (2006)
5. Raniwala, A., Gopalan, K., Chiueh, T.-C.: Centralized Channel Assignment and Routing Algorithms for Multi-channel Wireless Mesh Networks. *ACM SIGMOBILE Mobile Computing and Communications Review* 8(2), 50–65 (2004)
6. Raniwala, A., Chiueh, T.-C.: Architecture and algorithms for an IEEE 802. 11-based multi-channel wireless mesh network. In: *Proceedings of IEEE Inforcom 2005- 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, March 13-17, vol. 3, pp. 2223–2234 (2005)
7. De Couto, E., Aguayo, D., Bicket, J., Morris, R.: High-throughput path metric for multi-hop wireless routing. *Wireless Networks* 11(4), 419–434
8. Draves, R., Padhye, J., Zill, B.: Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks. In: *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, Philadelphia, PA, USA, pp. 114–128 (2004)
9. Subramanian, A.P., Buddhikot, M.M., Miller, S.: Interference aware routing in multi-radio wireless mesh networks. *Wireless Mesh Networks*, 55–63 (2006)
10. Yang, Y., Wang, J., Kravets, R.: Interference-aware Load Balancing for Multihop Wireless Networks. Tech. Rep. UIUCDCS-R-2005-2526, Department of Computer Science, University of Illinois at Urbana-Champaign (2005)
11. Yang, Y., Wang, J., Kravets, R.: Designing Routing Metrics for Mesh Networks. In: *WiMesh* (2005)
12. Sobrinho, J.L.: Algebra and Algorithms for QoS Path Computation and Hop-by-Hop Routing in the Internet. *IEEE/ACM Transactions on Networking* 10(4), 541–550 (2002)

# A Wireless Coordinating Scheme Proposed to Improve the Dispatching Security in China's Railway System

Junfeng Lu<sup>1</sup> and Wen-qiang Lu<sup>2</sup>

<sup>1</sup> Key Laboratory of Cryogenics, Technical Institute of Physics and Chemistry of Chinese Academy of Sciences, Beijing, China, 100190

<sup>2</sup> Graduate University of Chinese Academy of Sciences, Beijing, China, 100049  
junfenglu@mail.ipc.ac.cn

**Abstract.** In this paper, we proposed a wireless coordinating networking scheme along china's high-speed railway. By combing GPS coordinating system with WIFI (802.11n) networking system, a relatively secured design for railway dispatching is structured.

**Keywords:** GPS, WIFI, wireless dispatching, railway dispatching, 802.11n.

## 1 Introduction

In the construction program of Chinese high-speed railway system, concerns are mostly focused on the security issues in the dispatching system recently due to the unfortunate crash of bullet train in Wenzhou, China. According to the official explanation for the accident, the thunder strike which causes the wrong signal switch could be the major reason for the accident. However, from this one can also see that a frontier coordinating system with multiple wireless-communication links are badly needed by high-speed railway in China.

In today's china, railway system is still the major passenger transporting system between cities because the transporting fare is much chipper than that of airplane according to people's monthly payment (this is much different from western developed countries such as USA, where people travel by airplane or individual cars). Thus, to accelerate the cooperated development speed between cities in China, the advocating of high-speed railway system becomes one of the major tasks of the Ministry of Railway of the People's Republic of China (MOR). Today, China has the most advanced technologies applied to high-speed trains. However, the supporting facilities for this giant system are still under improvement, such as automatic dispatching system, digital network video surveillance security system, etc... According to the characteristics of the railway system, wireless related dispatching systems are the only choice for the communication between trains and command centers. And the system should be automatic and stable to avoid fatigue driving as well as careless dispatching mistakes produced by people. Thus, GPS related coordinating systems combined with wireless communicating systems apparently catch big eye-balls for the management of security related issues of today's railway systems.

Nowadays, GPS coordinating systems were applied in different security related areas such as: cab dispatching, fire fighting, children cell-phone, emergency rescue, etc... In general, contemporary GPS related vehicle control systems are mostly used by cab dispatching [1]. Some of them were designed on a paging system [2], as well, some of them, such as in [3], were interactive dispatching systems (or, it is not a real-time design). However, none of the above designs could be applied to a railway system. The design in [4] and an AVLDS (Automatic Vehicle Location and Dispatch System) mentioned in [5] was automatic real-time dispatching system, but their application in the railway system in China still needs more modifications. Moreover, the above cab dispatching system designs do not care about weather conditions which are severe to railway security.

In case of bad weather, sometimes GPS signals are very weak to be detected, thus, we still need other wireless facilities to coordinate the high speed trains. As the long distance wide-band wireless communication system, WIFI (802.11n) was successfully used in a bunch of applications in China, such as subway systems in some large cities of China. By overcoming the switching problems between WIFI access points, the rapid seamless roaming was approached by those applied systems. Thus, people can use their PCs as well as WIFI enabled phones conveniently inside the subway trains.

All in all, the design of a whole-weather sustainable dispatching system applied to Chinese high-speed railway system is still required. In the following of this paper, a GPS/WIFI based wireless dispatching scheme is discussed.

## 2 System Model

The system model of a wireless coordinating and communicating system for high-speed railway is more or less like a wireless cellular network model proposed in [6]. The difference between these two is that, the cells or access points (APs) of the high-speed railway system are paved along the whole pathway of the rail with a distance interval of  $D$  (typically this distance is around several hundred to several thousand meters according to 802.11n protocol) as shown in Fig. 1. In Fig. 1, as an example, a bullet train holding a GPS/DTD [1] (Data Transfer Device, it was defined as wireless communication link, such as GPRS, WIFI, etc.) runs on a rail with speed  $v$ , and communicates with outside world through WIFI APs. And these APs further communicate with Traffic Control (TC) centers [1] (through Ethernet, optical cable, or other sort of fast communication links) located in different railway section areas as shown in Fig. 2.



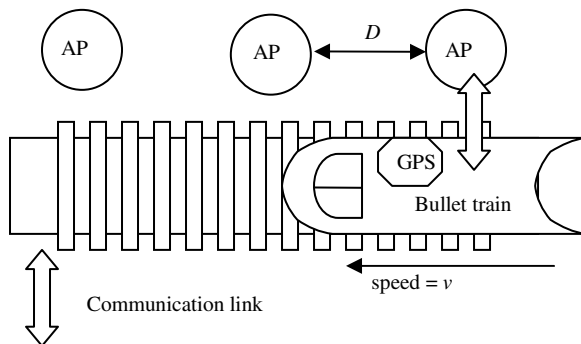


Fig. 1. System model

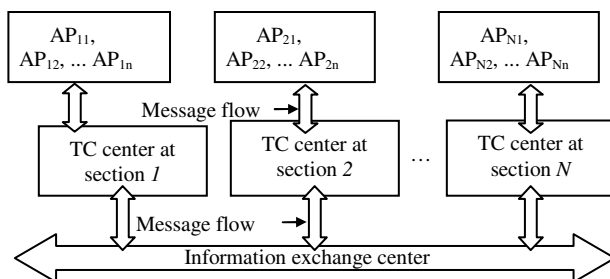


Fig. 2. Topology of traffic control centers

Holding the above topological model, the position of a bullet train can be precisely located. The whole process is explained in the following:

Regularly, a bullet train travels on the rail transferring its GPS location information to TC center  $i$  through GPS/DTD with a time interval;

1. Regularly, a bullet train travels on the rail transferring its GPS location information to TC center  $i$  through GPS/DTD with a time interval;
2. In case of weak GPS signal, DTD on the train will scan the best WIFI beacon signal and associate the strongest AP; meanwhile, DTD transfers its identification message (IDMSG), GPS weak message (GPSWKMSG), and current speed value  $v$  to TC center;
3. Normally, TC center receives the GPS position information of a bullet train through the GPS/DTD setup on the train, and listens to GPSWKMSG of any train through WIFI APs. It also maintains a position database (PD) that records the location of each AP registered inside its railway section area. So, in case of receiving a GPSWKMSG, the system will first identify the train by IDMSG, and locate it by searching the location of the associated AP inside its PD. Meanwhile, TC center calculates the next possible association AP for the train. Since the train is moving with a speed of  $v$ , the distance of next associating AP should be at least:

$$v * (\Delta T_{\text{Re-associate}} + \Delta T_{\text{disassociate}}) \tag{1}$$

4. TC center will then send back an IDMSG of next possible association AP (Note: this AP could be administrated by another TC center of railway section) to the train (the location of this AP is determined by Eq. (1)), and let the train prepare for re-association. (TC centers also use the real-time position information of every train on the rail for dispatching. Whenever there was a problem, it will continuously broadcast an error message (ERRMSG) to every train on the rail only through short distance WIFI communication link. Since the APs are deployed along the rail everywhere (the distance between any two APs is mostly within 1km according to 802.11n), so, a failure of one or two APs cannot influence an emergency stop of a bullet train.)
5. At the time of re-association, the DTD on the train will use the IDMSG sent from original TC center as a reference.
6. If the GPS signal comes back, the DTD will send to its current TC a "GPS signal back" (GPSBKMSG), and switch to step 1.

Obviously, one can also find that step 4, and 5 help the system to reduce switching time between WIFI APs (especially those APs located across different TC section areas).

### 3 Discussion

Normally, a bullet train exchanges its data through a DTD with TC centers, and WIFI system paved along the rail are used as internet service for on board passengers. Since the DTD uses GSM or 3G data transfer link, its roaming behavior is just like a cell phone (Fig. 3). However, when the signal of GPS is weak, the train cannot tell TC center its current location. Thus, the DTD on the train will switch its communication (with TC center) to WIFI link (Fig. 4). Meanwhile, the GPS/DTD still detects the GPS signal, whenever it's strong enough, the device switch its transfer back to normal. To explain this, as shown in Fig. 4, assuming that the DTD associates with AP<sub>1</sub> at time *T* and send GPSWKMSG as well as IDMSG with its speed value *v* to TC<sub>*i*</sub> when it loses GPS signal, TC<sub>*i*</sub> receives them, and then queries its PD for the precise location (*x<sub>1</sub>*, *y<sub>1</sub>*) of associated AP<sub>1</sub>. The value of (*x<sub>1</sub>* ± Δ*x*, *y<sub>1</sub>* ± Δ*y*) (Δ*x* and Δ*y* < *v* \* Δ*T<sub>association</sub>*) is the current location of the train. After locating the position of the train, TC<sub>*i*</sub> will calculate the best re-association AP<sub>1</sub> using Eq. (1) for the train to re-associate. Let:

$$D_i = v * (\Delta T_{\text{Re-associate}} + \Delta T_{\text{disassociate}}) \tag{2}$$

In most cases, *D<sub>i</sub>* is not a multiple value of *D*, and this distance value must be fallen somewhere between two APs, let's name these two APs as AP<sub>*i-1*</sub> and AP<sub>*i*</sub>. Since the train is moving ahead to AP<sub>*i*</sub>(Fig. 4), to provide more communication time, obviously, TC<sub>*i*</sub> will always choose further AP (AP<sub>*i*</sub>) as the next re-association point. And the address of this AP is sent back to the train by IDMSG. The train uses the address of AP<sub>*i*</sub> as a reference for re-association when roaming inside the WIFI system. Since the re-association AP was known ahead, the time of re-association process is obviously saved.

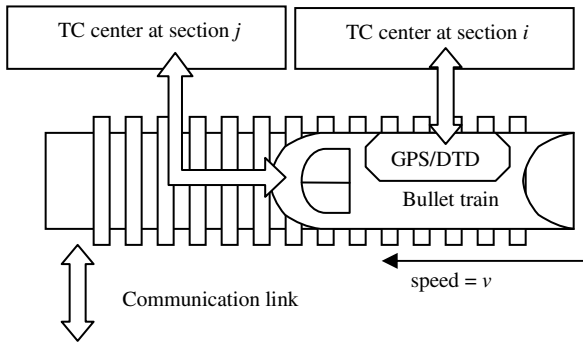


Fig. 3. Roaming between TC centers like a cell

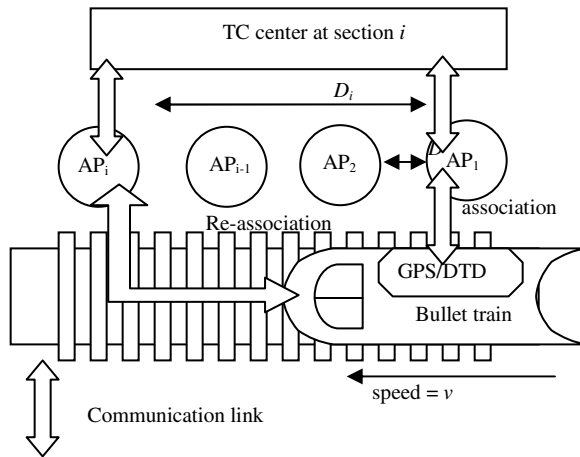


Fig. 4. Switching to WIFI system in case of weak GPS

Sometimes  $AP_i$  is not located in the area controlled by  $TC_i$  center as shown in Fig. 5. In this case, before the re-association, and to avoid an initial association and authentication between the bullet train and the  $TC_{i+1}$  center, the  $TC_i$  center will contact with  $TC_{i+1}$  center and transfer the identification information of the bullet train to  $TC_{i+1}$  through the information exchange center. After the  $TC_i$  center gets the authentication pass acquired from  $TC_{i+1}$ , it sends back the pass to the train for the seamless re-association between the train and the foreign  $TC_{i+1}$  center.

In case of an AP failure, the TC control center will always choose the next AP ahead as the re-association reference for the train. Since the successive AP failure is rare, the bullet train can definitely reports its location timely to avoid collision.

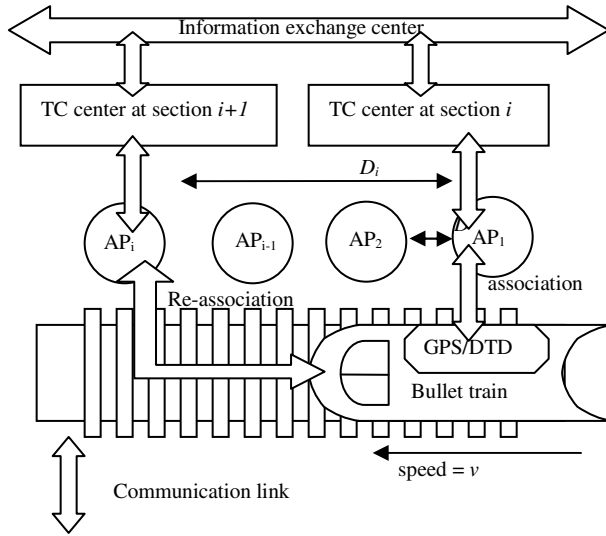


Fig. 5. Transfer to the next TC center

In China, the dispatching of a bullet train or a normal train is realized one by one between time intervals ( $\Delta T$ ). By adopting the system scheme proposed in this paper, a TC center of a specific railway section should have all the location information of each train on the rail. Whenever there is a problem or emergency stop of a train, the TC center will inform all the others on the rail to avoid crash. Obviously, the dispatching time interval  $\Delta T$  must at least satisfy the following condition:

$$\Delta T \geq \frac{D_{braking}}{v} \tag{3}$$

where  $D_{braking}$  is the braking distance of the train. Considering the association time and disassociation time of the train, Eq. (3) can be further improved as:

$$\Delta T \geq \frac{D_{braking}}{v} + \Delta T_{association} + \Delta T_{disassociation} \tag{4}$$

in case of an emergency stop of a train with communication problem, the TC center should wait for communication signal from the train for a time, say,  $\Delta T_{wait}$ , and if this time is passed, the TC center should default that the train has a problem. In this case, the dispatching interval calculated by Eq. (4) should be further modified as:

$$\Delta T \geq \Delta T_{wait} + \frac{D_{braking}}{v} + \Delta T_{association} + \Delta T_{disassociation} \tag{5}$$

from Eq. (5), one can further derive that the speed of the train can not exceed:

$$v \leq \frac{D_{braking}}{\Delta T - \Delta T_{wait} - \Delta T_{association} - \Delta T_{disassociation}} \tag{6}$$

## 4 Conclusion

By analyzing the contemporary high-speed railway system in china, a relatively secured wireless coordinating and dispatching system scheme, which adopts a combination of GPS/DTD with WIFI, is practically proposed. Its coordinating and roaming features are detailed in the paper. And the dispatching mechanism is also discussed at the end of the paper.

**Acknowledgments.** This work is partially supported by NSFC Grant No. 51078378, 51176200, and 50936006.

## References

1. Lu, J.F.: A discussion on an efficient virtual traffic control system for un-trackable city vehicles. In: IEEE 2010 International Conference on Power and Energy Systems (ICPES 2010), pp. 29–32 (2010)
2. Xi, J.T., Xiao, H., Chen, Z.X., Wang, J.: Satellite automatic taxi dispatching and coordinating system. Chinese Patent, No.: CN01107414.0 (2002)
3. Umeda, N., Oyama, K.: Taxi dispatching system and dispatching method. US Patent Application Publication, Pub. No.: US 2006/0034201 A1
4. Magee, P.: Service ordering system and method. US Patent Application Publication, Pub. No.: US 2004/0153325 A1
5. Liao, Z.Q.: Real-time taxi dispatching using Global Positioning Systems. Communications of the ACM 46(5), 81–83 (2003)
6. Lu, J.F.: A History Statistical Estimation Method for Hand-off Probability in Wireless Cellular System. In: IEEE The 2nd International Conference on Industrial and Information Systems, IIS 2010 (2010), doi:10.1109/INDUSIS.2010.5565678

# The Prospects and Development on FTTH Technology in China

Jin Cai<sup>1</sup>, Zhiyong Tao<sup>2</sup>, Huijuan Zhang<sup>2</sup>, Lu Qin<sup>3</sup>, and Qian Mao<sup>2</sup>

<sup>1</sup> Hubei university of education

<sup>2</sup> Wuhan research institute of post and telecommunication

<sup>3</sup> Wuhan university

**Abstract.** FTTH has become a new generation optical network access technology along with the popularity of the application of digital television, IPTV and other high bandwidth multimedia applications. This paper mainly introduces the several stages in the development of FTTH in China, several main technologies of FTTH, and the formulation of related international and domestic standards. Finally, it'll discuss the prospects and development of next-generation FTTH technology.

**Keywords:** EPON, FTTH, network convergence, DSL, IPTV.

To realize the end-to-end optical network is a dream coming along with the initial application of optical communication. In the past 30 years, FTTH roughly experienced three stages of development since the concept first appeared. The first stage was in the late 1970s, from the first FTTH field trial of France, Canada and Japan. The second stage was the research and experiment of BPON in the United States and Japan around 1995. Those two opportunities were aborted because of the high cost and the lack of market demand. The concept of GPON and EPON are proposed and started being standardized from 2000, when FTTH market began to blossom, and when the global optical fiber access market grew rapidly. The number of global FTTH subscribers also shot up, the proportion of which would gradually increase in global broadband access. China, however, has started the FTTH research in the mid 1990s. So far, it roughly experienced several stages: such as technical selection, determining technology standard for pilot project, large scale application and self-innovative development.

## 1 The Stage of FTTH Technology Selection and Exploration

FTTH is a kind of access technology which completely adopts the fiber lines from central office (CO) to the user side. FTTH mainly includes point-to-point active optical access technology and point to multi-point passive optical network (PON) technology. A formal proposal G.983.1 was issued by ITU - T in 1998, since when the standards work of PON system based on ATM technology began. In 2001 APON changed its name to BPON. ITU - T issued the G.983.10 in 2004, which is the symbol of the completion of G.983 BPON standard series.

In 1995, Wuhan Research Institute of Post and Telecommunication (WRI) started to develop narrowband PON system, and the commercial application was realized. The main technical characteristics were to provide the TDM services for PSTN with point to the multi-point topology. In 1999, this first generation narrowband PON communication products from Fiberhome had been deployed in Guangxi and Liaoning provinces. But because of the complicated technology and high cost, it was not applied widely.

Until March of 2001, the full service access system of state “863” planning project was completed by Wuhan research institute of post and telecommunication (WRI), namely APON research projects, and was approved up to standards by “863 project” experts. In 2002, Fiberhome and Huawei successively developed related products. But because of the stagnation of ATM network construction. The APON technology was seldom used in China.

This stage is the period of network technology transformation with the rapid development in optical fiber communication field. All sorts of point-to-point and point-multipoint technologies were successively applied to fiber-optic access network. The high degree of standardization could help to save the OLT (Optical Line Terminal) optical interface and fiber, improve the system’s expansibility, and bring convenience to maintenance and management. Although the deployment of the network ATM technology was not as good as expected, and BPON product has never been widely used, it’s clear that the PON will be the main mode of realization of FTTH. BPON now is the foundation of broadband PON technology. All other PON technologies which were later developed have directly or indirectly adopted most of the standard of BPON series.

## **2 The Stage of China's FTTH Standards Determination and Pilot Project Deployment**

In 2000, the first mile Ethernet working group (EFM, 802.3 ah) was established by IEEE. The EFM proposed the PON technical standard based on Ethernet. This is what we are familiar with- EPON standards. In 2004, IEEE802.3 ah EPON protocol was formally approved. EPON has combined the low-cost Ethernet and PON’s structural characteristic, which gained a wide acceptance in its own industry even before the formal standard was launched. EPON has simpler encapsulation and more efficient compared with the traditional BPON. Its rate was also promoted to 1Gb/s. EPON enhanced its characteristics in OAM compared with the conventional P2P Ethernet. The deployment costs and maintenance costs of EPON equipment will also be much lower than P2P fiber Ethernet.

GPON, developed based on APON/BPON, is a PON systematic standard whose transfer rate is more than 1Gb/s. It was proposed by Full Service Access Network Forum (FSAN) organization in 2001. GPON standards G.984.1 and G.984.2 were formally approved by ITU-T in 2003. G.984.3 and G.984.4 were approved in 2004 one after another. Then G.984.x series of standards was then formed. After that, G.984.5 and G.984.6 were launched, and the increased bandwidth and distance extension of GPON are defined respectively in them.

EPON products were launched early, but they had no mature standard for reference. The national "863" project "based on gigabit Ethernet broadband passive optical network (EPON)" experiment system was accomplished by Fiberhome from 2002 to 2003. In 2005, the China EPON pilot project was deployed in Zi-Song garden by Fiberhome and Wuhan Telecom Company. This successful project was the product of national "863" program which could provide users with the service of ordinary telephone, fax, broadband Internet access, CATV, IPTV and so on. This project symbolized the commercial use of FTTH in China. The development of FTTH was supported actively by Wuhan government where China's optical valley is located. They have done a lot of work, such as local policy publication, technology standard formulating, user/real estate enterprise/operation /manufacturing enterprises cooperation, government reorganization and so on. They completed a series of FTTH local standards in Wuhan to better promote the development of FTTH in 2005.

At the same time, the attention in access network was given to FTTH by China's telecom operators. The chip-level EPON interoperability testing was promoted in 2005. The EPON system-level interoperability testing was conducted in 2006. Then in 2007 the enterprise standards V1.3 was published according to these device assessment test. The EPON equipment, large-scale chip and system-level interoperability was realized by China Telecom for the first time internationally. The pilot projects were started from Wuhan, Beijing, Shanghai, Guangzhou and other cities. Then, a lot of FTTH pilot projects were deployed in all of the provinces (autonomous regions) in China from 2006 to 2007. So the construction cost for FTTH has dropped substantially for a large-scale application conditions, which indicates that China FTTH technology standards have been identified, and the large-scale application has become possible after the pilot project testing.

### **3 The Stage for FTTH Large-Scale Application and Innovation Development in China**

In 2005, a number of important FTTH industry standards and research programmers were completed by China Communications Standards Association (CCSA) organizations in the 3rd plenary meeting of Transport and Access Network Technical Committee (TC6). The draft of "broadband optical access network and FTTH architecture and general requirements" was published by Fiberhome. In the drafting process, the FTTH system architecture, network topology, type of services, technology and implementation requirements, performance requirements, and operation and maintenance requirements etc. are included. The FTTH optical cables and specifications for ancillary facilities were also outlined. The "Broadband Optical Access Network and FTTH architecture and general requirements of" provides a reference and guidance for the future standards work. "Access network technical requirements - Based on Ethernet Passive Optical Network E-PON" (YD/T1475) and "Access network equipment, test methods - Based on Ethernet Passive Optical Network EPON" (YD/ T1531) became the industry standard in 2006. Under the leadership of China Communications Standards Association (CCSA), the relevant committees and working groups have carried out a lot of FTTH-related standardization work. The FTTH industry standards covering the active / passive



components, systems, fiber optic cable and other aspects have been completed in recent years.

In 2007, China Telecom held a spot meeting in Wuhan. Leaders of China telecommunication planning and construction from 10 provinces visited Fiberhome FTTH pilot project for a variety of scenarios. The "optic fiber replacing copper pairs" strategy was mainly discussed for the future of China Telecom access network. The first construction and deployment Central Purchasing amounted to 2 million in 2008. China Unicom also conducted the pilot, testing, selection and Central Purchasing, investing ¥15.0 billion in the implementation of FTTH in some areas. China Mobile, radio and television sectors and other private networks have also taken appropriate actions to promote the various applications of FTTH.

Fiberhome, ZTE and Huawei's EPON products have been put on the market, widely used by the operators and network users all over China from 2004 to 2008. The networked capacity goes above 10 million. EPON products, practically tested by scale-commercial use, have been spread in more than 30 provinces.

"The China optical access Industry Alliance" was officially established in Beijing on October 24, 2008. The first participants include telecom operators, telecom equipment manufacturers, manufacturers of optoelectronic devices, optical fiber and cable manufacturers, telecommunications and design units, research institutes and universities and so on. According to the league constitution, the Alliance Secretariat is located in Wuhan Institute of Posts and Telecommunications in China Optical Valley. "The China optical access Industry Alliance" is formed to actively promote the fiber access, including fiber to the home of technological progress and industrial development, to promote the development of China's communication industry and improve the national information and communications innovation and core competitiveness. "The China optical access Industry Alliance" can take greater advantage of domestic institutions in the field of optical access. The Alliance has formed a good basis to actively explore various forms to enhance the fiber access link chain industrial cooperation, technical cooperation and knowledge sharing, to improve the level of fiber access technology innovation and competitiveness, and to promote industrial development. The establishment of the League symbolized that China's FTTH technology has entered a new phase of scale production and innovative development.

#### **4 FTTH Technology Development Prospects**

In the long term, the existing bandwidth and split ratio of FTTH technology are still unable to meet the bandwidth need of 50M ~ 100Mb/s per user for the future development. So the Next generation PON has become the research hotspot now. Both EPON and GPON will follow the direction of those of a higher rate as 10G. Currently the international standard organization has already began the discussion and formulation of the standards. The symmetrical rate 10G EPON standard has already been finished in September, 2009, and some of the asymmetrical rate 10G EPON system has been put into use online. In September, 2009, 10G EPON ITU has standardized the physical layer, and was estimated to complete the process of standardization until 2011. ITU-T and FSAN separate NG-PON standard into two

phases: NG-PON1 and NG-PON2, the former of which is orientated as an advanced technology in the metaphase research, covering the years 2009 to 2012, and the latter is a long-term solution, covering the years 2012 to 2015. Since some of the basic technologies such as PMD and TC layer are still in the phase of technique demonstration, which somehow restrains the industry chain development, NG-PON1 standard was started late. The development of optical module and chip industry is similar to that of GPON. If the index parameters defined by NG-PON standard, such as on/off time of burst module and synchronous sequence are still too precise and accurate, the maturity of scale-commercial use of device industry will become more difficult, which will consequently postpone the process of industry chain.

The special wavelength is assigned to each ONU in WDM PON technology. WDM PON technology can transport all the traffic of various protocols transparently to meet the bandwidth needs for a long time. The national "Eleventh Five-Year" 863 major projects, "multi-wavelength ( $\lambda$ -EMD) low-cost Ethernet integrated access system" was completed by Fiberhome together with China Telecom and other units in 2010. They developed the WDM- TDM PON system successfully. The WDM- TDM PON system provides 32 waves in a single-fiber, and supports 1:64 splitting ratio for each wave; the transmission distance is up to 20 kilometers. The mixed mode of Using WDM and TDM PON structure can be compatible with existing 1G/2.5G/10G EPON, GPON and P2P, and other fiber access technology. So this WDM- TDM PON system is facilitated to the realization of "triple play" services to realize next-generation optical access technology by carrying the existing CATV services.

Increasing bandwidth demand will be driven by new services; operators must consider providing higher bandwidth and operational capacity to meet growing demand and competitive pressures in the construction of PON network. So the next-generation PON technology will be evolved in several directions as follows:

1. To be a higher rate: 10G EPON / GPON interface provide 10 times bandwidth of existing EPON / GPON. So the next generation of PON technology increases the user bandwidth from 50Mb/s to 100Mb/s, and the number of users from 256 to 512. The cost of each user would be reduced.

2. To be a longer distance and more splitting ratio: the optical power budget of next generation PON should be greater than 28dB, and the split ratio should be from 32 to 128 or even higher. To reduce the number of Bureau and the cost of operation and maintenance by extending the range from 20 to 60 km.

3. To be more wavelengths: The improvement of WDM-PON can also realize the transparency of agreements and a higher user rate up to 100M or even 1G.

The next 5 years, the capacity of new construction of optical access is expected to be more than 80 million lines. The fiber to the home will not only be the main method to achieve the "triple play" universal broadband network and enhance the level of IT application / social informationization for China, but also gradually develop into a leading role in emerging industries of hundreds billion Yuan, which will definitely spur the national economic growth.

## References

1. ITU-T Draft Recommendation G.988 (ex G.omci) (new) ONU management and control interface (OMCI) specification
2. Mao, Q.: FTTH Application Coming to a Climax in China. Designing. Techniques of Posts and Telecommunications, pp.1–5 (June 2008)
3. Tanaka, K., Agata, A., Horiuchi, Y.: KDDICorp., Tokyo, Japan. IEEE 802.3av 10G-EPON Standardization and Its Research and Development Status, February 15, pp. 651–661 (2010)
4. Shen, C.-B.: Progress & Application of Next Generation PON Technology, pp. 1–8 (August 2010)
5. Xiao, L.: A Novel WDM-PON Structure Using the Orthogonal FSK/ASK Re-modulation Scheme. In: Proceedings of OSA/ACP (2009)

# Cooperative Behavior in Evolutionary Snowdrift Game on a Growing Complex Network

Qingke Zeng<sup>1,\*</sup>, Haiyong Yang<sup>2</sup>, and Hong Wang<sup>1</sup>

<sup>1</sup> School of Science, Southwest University of Science and Technology,  
Mianyang, China, 621002

kewhke@163.com

<sup>2</sup> China Aerodynamics Research & Development Center,  
Mianyang, China, 621000

**Abstract.** In this paper, we study the evolutionary snowdrift game on a growing complex network with acceleratingly increasing number of nodes. In this growing network, the number of nodes added at each time step is a function  $c \times n(t)$ , where  $n(t)$  is the number of nodes present at time  $t$  and  $c$  is a tunable parameter. The evolutionary SG plays with a preferential selection probability proportional to  $\exp(U_j/k_j) * A$  and a new strategy updating probability  $H_{i \rightarrow j} = 1/(1 + \exp[(U_i/k_i - U_j/k_j)/k])$ , where  $U_i$  denotes the total payoff of player  $i$ ,  $k_i$  denotes the degree of  $i$ ,  $A$  and  $k$  are tunable parameters. It is shown that the cooperation frequency depends only on the payoff parameter  $r$ , it is much enhanced in the range of small payoff parameters, but descends straightly with the parameter  $r$  increasing, and it has nothing to do with the parameter  $c$  and  $A$  with this mechanism.

**Keywords:** Snowdrift Game, Simulation, Complex Network, Cooperation Frequency.

## 1 Introduction

In nature, systems consisting of competing entities such as physical systems, biological systems, economic systems and social systems are ubiquitous [1]. Contrary to the context of Darwinian's evolution, cooperative behavior plays an important role in these systems. To understand how cooperation emerges and persists in systems consisting of competing entities, scientists often resort to Evolutionary Games Theory, which has been considered a common mathematical framework for investigating the emergence of cooperative behavior in a competitive setting[2,3]. The Evolutionary Games Theory developed from classical game theory, in which the agents are rational and the game is solid. The Evolutionary Games Theory modified these two shortcomings and supposed the game to be evolving and the agents' rationality are bounded. But the Evolutionary Games can't give a steady cooperative state. As a

---

\* Corresponding author.

typical extension, the spatial game is introduced by Nowak and May [4] and it can result in emergence and persistence of cooperation in PD. Motivated by their work, Santos and Pacheco[5] discovered that scale-free networks provide a unified framework for the emergence of cooperation . In contrast, Hauert and Doebeli [6] found that cooperation is inhibited by the spatial structure.

The evolutionary prisoner’s dilemma (PD) and the evolutionary snowdrift game (SG) have drawn much attention from scientific communities. In PD and SG games, agents play the games in pairs with their direct neighbors. Each agent has to decide simultaneously whether it wishes to cooperate (C) or to defect (D). If both of them choose to cooperate, the both get rewards R, and P for mutual defection. While if one cooperates and the other defects, the cooperator gets S and the defector gets T. The payoff matrix can be described as

$$\begin{matrix} & C & D \\ C & (R & S) \\ D & (T & P) \end{matrix}$$

In PD, the rank of these elements is  $T > R > P > S$  and  $2R > T + S$ . Thus, if the neighbor chooses C, the agent prefers D to C (for  $T > R$ ), and if the neighbor chooses D, the agent still chooses D since  $P > S$ . That means the best choice of the agent is always D. So the PD constrains the emergence of cooperation. Then, the proposal of snowdrift game (SG) was generated to be an alternative to the PD [7]. The only difference between PD and SG is the rank of elements. In SG,  $T > R > S > P$ , the best strategy of the agent depends on its neighbors. If the neighbor chooses C, the agent chooses D since  $T > R$ . If the neighbor chooses D, the best choice is C (for  $S > P$ ). In this way, the cooperation emerges more easily in SG than PD. Without losing generality, we can set  $R=1, S=1-r, T=1+r$  and  $P=0$ , where  $0 \leq r \leq 1$  indicates the rate of labor cost [8]. The SG payoff matrix is

$$\begin{matrix} & C & D \\ C & (1 & 1-r) \\ D & (1+r & 0) \end{matrix}$$

Many works have been done on the snowdrift game [5,8-11]. Previous works described the cooperation behavior within the following context. One is how the agents revise their strategies. Many evolution rules, such as memory, learning strategy etc have been discussed. It is found that memory and learning strategy promote cooperation behavior [8,9].The others pay much attention to different population structures, including regular networks and growing complex networks. Three types of networks are most usually discussed. They are the ‘Regular Network’, the ‘Small-World Network’ and the ‘BA Network’. Most of the simulations showed that spatial structure promotes cooperation, but spatial structure is not necessarily beneficial in promoting cooperation in the SG[6].

BA Network was generated by Barabási and Albert[12]. The BA network is grown from a few nodes and one new node is added at each time step. The new node links to  $m$  existing nodes with the probability being proportional to their degrees. It is a wonderful model to describe the real world as a system with power law degree

distribution. In the BA Network, only one node is added at each time step, but the increasing of nodes of Internet is accelerating [13].

In order to describe the real world more clearly, Li Ji *et al.* modified the BA Network to a ‘Growing complex network model with acceleratingly increasing number of nodes’[13], in which the nodes added at each time step is  $c \times n(t)$ , where  $n(t)$  is the number of nodes present at time t, c is a tunable parameter. Due to the importance of the spatial structure, it is useful to study the evolution games on the modified network. Therefore, in the present work, we investigate the evolutionary SG on the modified network.

This paper is organized as follows. In the next section, the model of evolutionary SG and the learning rule is described. The simulation results and discussions are given in section 3. Finally the paper is concluded by the last section.

## 2 The Model

We consider the evolutionary SG with agents located on the growing complex network with acceleratingly increasing number of nodes. The network is constructed by the following rules: (1) starting from  $m_0 = 4$  fully connected nodes; (2) at each time step,  $c \times n(t)$  nodes is added; (3) The number of new links added with each new node is  $m = 4(m \leq m_0)$ . Especially, the new nodes will not attach to the other new nodes [13]. The existing node  $i$  is chosen to be connected with the attachment probability given by

$$P_i = \frac{k_i}{\sum_l k_l},$$

where  $k_i$  is the degree of node  $i$  which denotes the numbers of nodes attached to  $i$  and  $l$  runs over all the existing nodes[12]. That means the nodes with higher degree have more probability to be attached and it is based on the truth that people with more prestige have more influence in the society.

Initially, the strategy of C and D is randomly distributed on the nodes in the growing complex networks and each percent of them is 50%. At each time step, the agents play SG with their direct neighbors simultaneously and get payoffs depending on the payoff matrix. The total payoff of player  $i$  at each time step is stored as  $U_i$ . When play  $i$  is updated, it firstly chooses a neighbor  $j$  to learn from with the probability

$$P_{ij} = \frac{\exp[(U_j/k_j) * A]}{\sum_{l \in \Omega_i} \exp[(U_l/k_l) * A]},$$

where  $\Omega_i$  is all the numbers of direct neighbors of  $i$ , A ( $A > 0$ ) is a tunable parameter.

The probability above is modified from  $P_{ij} = \frac{\exp(U_j * A)}{\sum_{l \in \Omega_i} \exp(U_l * A)}$  [14], since the normalized payoff will describe the benefit of agents’ strategy more clearly than the total payoff.

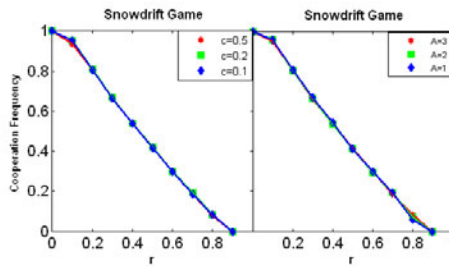
For  $A > 0$ , the agent will learn more probably from the neighbor with higher payoff. Agent  $i$  adopts  $j$ ' strategy as its new strategy with the probability

$$H_{i \rightarrow j} = \frac{1}{1 + \exp[(U_i/k_i - U_j/k_j)/k]}$$

where  $k (0 \leq k \leq \infty)$  is a parameter denotes the amplitude noise.. In this paper, following the previous works,  $k = 0.1$  [8].

### 3 Simulation and Analysis

Simulations were carried out for  $N=10000$  agents located on the complex network at least. We mainly concerned on the influence of every parameter on the cooperation frequency  $f_c = \frac{N_c}{N_c + N_d}$ , where  $N_c$  denotes the numbers of agents who choose C as their strategy and  $N_d$  denotes the numbers of defectors. Obviously,  $f_c$  ranges from 0 to 1, where 0 and 1 correspond to the ALL D state and ALL C state respectively. Firstly, we set  $A=3$  and study the cooperation frequency  $f_c$  for  $c=0.5, c=0.2, c=0.1$ , respectively. Secondly, we set  $c=0.1$  and study the  $f_c$  for  $A=3, A=2$  and  $A=1$  respectively. Results are calculated after the system reaches equilibrium state. Each data point results from an average of over 20 realizations. The results are shown in the fig 1.

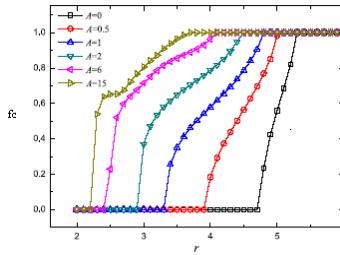


**Fig. 1.** The cooperation frequency as a function of  $r$  for different values of  $c$  &  $A$

The left of fig.1 shows  $f_c$  as a function of  $r$  for different values of  $c$ . The main features of  $f_c$  are:(1) The value of  $f_c$  descends from 1 to 0 as  $r$  increases from 0 to 0.9. Thus, the system reaches to a AllC state when  $r=0$  and a AllD state when  $r \geq 0.9$ . This confirmed that people prefer D to C when the labor cost is too large.(2) With  $A=3$ , the value of  $f_c$  has nothing to do with  $c$ , it depends only on the payoff parameter  $r$ . That means the  $f_c$  will not change when the network's growing speed is different.

The right of fig.1 shows  $f_c$  as a function of  $r$  for different values of  $A$ . The main features of  $f_c$  are quite similar to the left of fig.1. They are : the value of  $f_c$  descends as  $r$  increases from 0 to 0.9 and it has nothing to do with  $A$  for the fixed  $c=0.1$ . That means when the growing speed of network is fixed, the cooperation frequency depends only on the payoff parameter  $r$ .

The results in SG above are quite different from the results in the public goods game with the preferential selection  $P_{ij} = \frac{\exp(U_j * A)}{\sum_{l \in \Omega_i} \exp(U_l * A)}$ . The results in public goods game are shown in Fig 2[14]:



**Fig. 2.** The cooperation frequency  $f_c$  as a function of  $r$  for different values of  $A$  in public goods game

In Fig2, the preferential selection promotes cooperation over a wide range of  $r$ , the  $f_c$  increases as the parameter  $A$  increases. But in SG, the cooperation frequency has nothing to do with  $A$  and the cooperation frequency descends quickly as  $r$  increasing. That is, the cooperation behavior isn't stable. The reason why the preferential selection doesn't promote cooperation can be understood: higher payoffs aren't the only goal in a system evolution.

### 4 Conclusion

In summary, we have studied the evolutionary snowdrift game (SG) on a growing complex network with a preferential selection probability proportional to  $\exp[(U_j/k_j) * A]$  and a new strategy updating probability  $H_{i \rightarrow j} = 1/(1 + \exp[(U_i/k_i - U_j/k_j)/k])$ . The results show that the cooperation frequency depends only on the payoff parameter  $r$ . It is much enhanced in the range of small payoff parameters, but descends straightly with the parameter  $r$  increasing. The results are quite different from the results of the public goods game, in which the preferential selection promotes cooperation obviously. And the result confirmed the view that the spatial structure often inhibits the evolution of cooperation in the snowdrift game [6].

**Acknowledgments.** The authors would like to thank the Doctor Foundation of Southwest University of Science and Technology under Grant No 10zx7138 for supports.



## References

1. Hammerstein, P.: *Genetic and Cultural Evolution of Cooperation*. MIT, Cambridge (2003)
2. Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge University Press, Cambridge (1982)
3. Gintis, H.: *Game Theory Evolving*. Princeton University, Princeton (2000)
4. Nowak, M., Sigmund, K.: Tit-for-tat in heterogeneous populations. *Nature (London)* 355, 250–253 (1992)
5. Santos, F.C., Pacheco, J.M.: Scale-Free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. E* 95, 098104 (2005)
6. Hauert, C., Doebeli, M.: Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature (London)* 428, 643–646 (2004)
7. Sugden, R.: *The Economics of Rights, Cooperation and Welfare*. Blackwell, Oxford (1986)
8. Du, W.-B., Cao, X.-B., Zhao, L., Hu, M.-B.: Evolutionary games on scale-free networks with a preferential selection mechanism. *Phys. A* 388, 4509–4514 (2009)
9. Zhang, M.F., Wang, B.H., Wang, W.X.: Randomness Effect on Cooperation in Memory-Based Snowdrift Game. *Chin. Phys. Lett.* 4, 1494–1497 (2008)
10. Wang, W.X., Ren, J., Chen, G.R., Wang, B.H.: Memory-based snowdrift game on networks. *Phys. Rev. E* 74, 056113 (2006)
11. Ni, Y.C., Xu, C., Hui, P.M., Johnson, N.F.: Cooperative behavior in evolutionary snowdrift game with bounded rationality. *Phys. A* 388, 4856–4862 (2009)
12. Barabási, A.-L., Albert, R.: Emergence of Scaling in Random Networks. *Science* 286, 509–512 (1999)
13. Li, J., Wang, B.H., Jiang, P.Q., Zhou, T., Wang, W.X.: Growing complex network model with acceleratingly increasing number of nodes. *Chin. Phys. Soc* 8, 4051–4056 (2006)
14. Shi, D.M., Yang, H.X., Hu, M.B., Du, W.B., Wang, B.H.: Preferential selection promotes cooperation in a spatial public goods game. *Phys. A* 388, 4646–4650 (2009)

# Research and Implementation on a Collaborative Style of Test Bank System

Yi-Bing Liu

Department of Mechanical and Electric, Shaoyang Professional-Technology College,  
Shaoyang Hunan 422000, China

**Abstract.** To develop a collaborative item bank system and implementation methods used in data communications and data processing devices. The test bank system includes: database module, test module, communication module and access module. The implementation of the system includes the establishment of test database, paper library, the table of assessment and the table of features. Periodic testing and feedback results are required. Online communication and research discussions are required according to results based on feedback. Apply the implementation of characteristics and retrieval to the measured object Construct and use this system according to the characteristics of the object . The bank system and the implementation can help online team-based teaching model to guide the students to learning and formative assessment.

**Keywords:** test bank systems, research, implementation.

## 1 Introduction

In order to assess students' knowledge fairly in all aspects, most of the universities conducted exams using test paper exam at present, and the system has been widely used. However, most schools only have the stand-alone offline test database with low frequency of its use and long update cycle, and because of personal reasons, they can not guarantee that the papers of the knowledge points and the difficulty of distribution meet the requirements of examinations, and nor can they ensure the fairness of the examinations. In addition, there are difficulties in the auditing, sorting and construction of the test database as a result of using the traditional time-consuming and labor-intensive manual method and inefficient, timeliness coordination. Therefore, it is difficult to ensure its quality. To overcome these shortcomings, we studied a collaborative exam system with its implementation methods. This system provides students, teachers, teaching director and department director with communication channels and a management platform. According to real-time characteristics of the test object, we work together dynamically to construct and use the test database in efficient ways to help online team-based teaching model to guide the students to learning and formative assessment.

## 2 The Bank System and the Implementation

The Q system uses a browser / server (B / S) as its operating mode, a system of a three-tier structure, which is of the middle of the front-end Web performance logic

and business logic from the ASP. NET driver, the back-end SQL server 2000 database using ADO. NET access.

### 2.1 Structure Module Exam System

The bank system structure shown in Figure 1

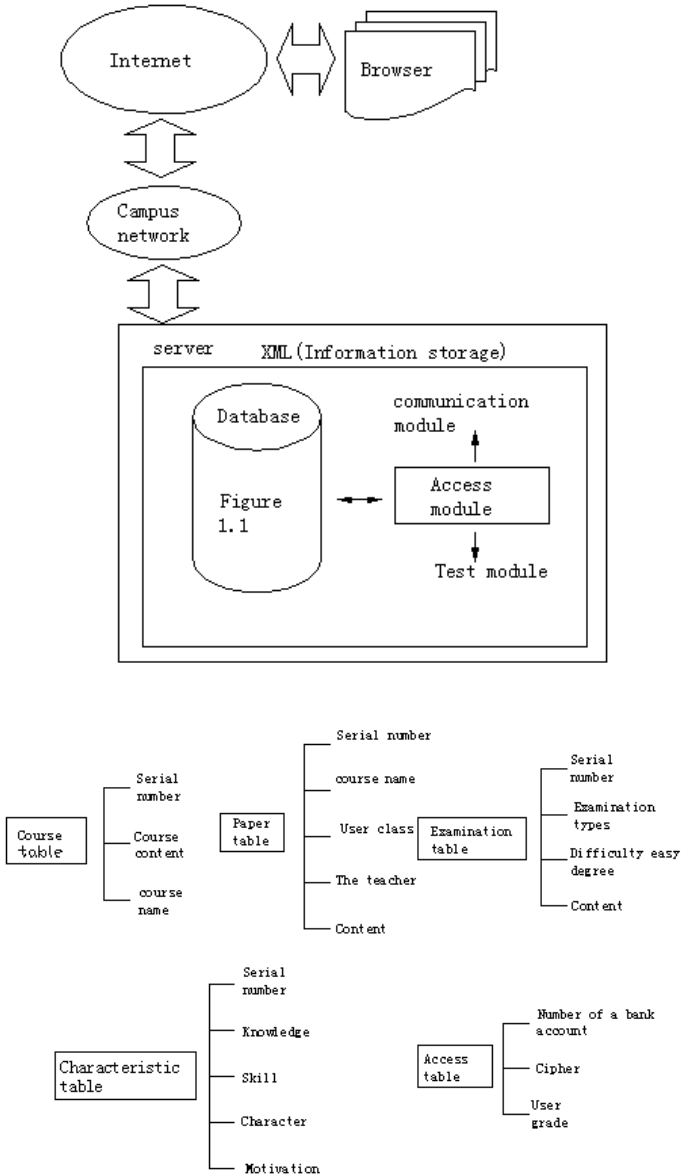


Fig. 1. Schematic diagram of exam system

(1) Access module: for each category and class of users to establish the corresponding data tables, and the identity of a particular region to manage user authorization control Used to redirect the solution to different user with different interfaces, the method principle is based on the login screen to determine the return address, then redirected to a different page. Write down an Admin use a page base class, unified in the base class for permissions page processing.

(2) Database module: ① course management modules: each point of the curriculum structure and course knowledge attributes (including the kinds of questions, difficulty, requirements, paper type, paper template); ② paper management module: manual and automatic test paper test paper (including syllabus and question type, degree of difficulty, automatically extracting questions); ③ questions management modules: input questions, import questions, test statistics.

(3) test module: According to the characteristics of identification papers taken from the paper library. Test consists of online testing and regular testing (paper-based exam papers, training experiments), the time monitoring the implementation of online testing, candidates should carry out an assignment in a specified time.

(4) communication module: The new embedded IM (instant messaging) tools, a PC client, mobile phone or WAP client to log on the exchange of information, exchange of content, including text, interface, voice, video and documents sent each other. — to provide status information display contact list features, contact person is in line with the contact person with the ability to talk.

## 2.2 Implementation

The bank system's implementation in the implementation of the flow chart shown in Figure 2:

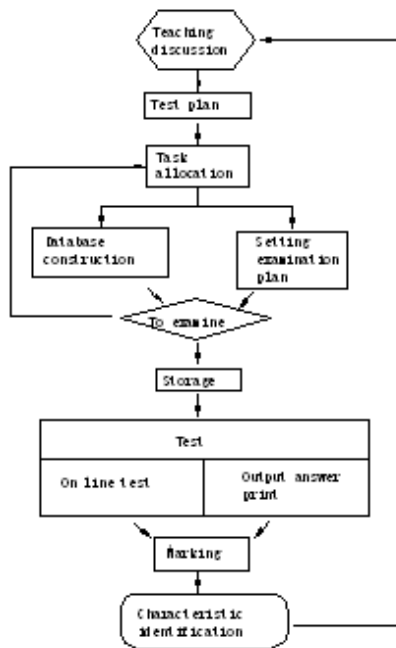


Fig. 2. Bank system execution flow chart

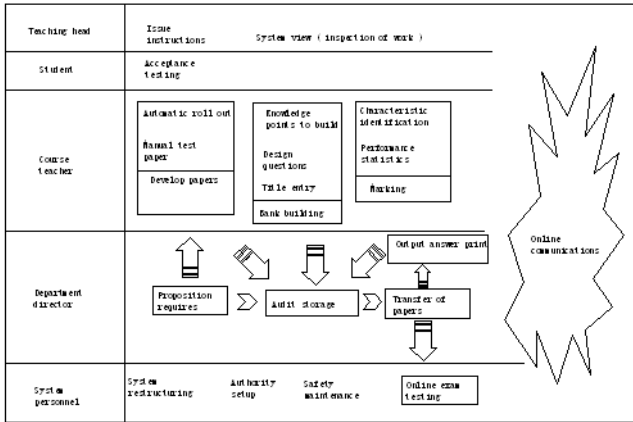


Fig. 3. The specific item bank system implementation diagram

The exam system is specifically related to implementation shown in Figure 3:

Teaching Director: work inspection (system view), issued an order;

Department Director: presided over the discussion of teaching and research, assignments, test paper to determine the program, test, examination paper storage, test, papers of the call;

Teacher: teaching and research involved in the discussion, test database construction, paper making, scoring, analysis of test results and test characteristics of the object identifier;

Students: participate in the discussion of teaching and research, tested;

System Administrator: to maintain system uptime.

### 2.3 Implementation of the Test Item Bank System Characteristics of the Object Identification Rules

Knowledge structures: classified according to the course content segments, as is said in dgjc1.3 "electrical-based" Chapter 1, Section 3;

Skills: each divided into three levels: good average and poor according to movement speed, accuracy, agility, coordination, stability and subdivision, ;

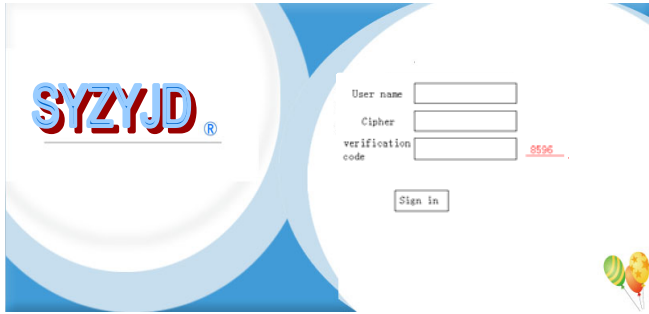
Personality: outgoing, introverted, neutral;

Motivation to learn: good, average and poor.

Apply the check box control to the knowledge structure , box control to write the remaining features of a single identification page. Characteristics are identified and confirmed by the necessary testing and evaluation of the teaching team. Manual identification and automatic search provide a reference for the test paper program to promote the positive transformation.

### 3 The Specific Implementation Modalities

Create data tables for users of different categories and levels , and manage and control uses in a particular region and with a particular identity by authorization. (Figure 4) Solve the problem for different users to log on different interfaces by redirecting, the method principle of which is to judge according to the returning address of the login interface, and then redirect to a different page.



**Fig. 4.** The user login screen

Create the folders named Admin and User on the site, which are meant to store files for admin and ordinary users respectively. You can also just set up a Admin folder. The method is to determine a returning path, therefore, folders used by admin users should be distinguished by the path. When accessing the admin folder, it will go directly to the interface of AdminLogin.aspx. After logging in, you can get the current user name as well as its roles using the following method on page in / Admin / folder, and then determine whether the current user has the authority to operate according to its role.

```
Response.Write("<li> Currently logged on user = " + Page.User.Identity.Name);
Response.Write("<li>admin = " + Page.User.IsInRole("admin"));
Response.Write("<li>reader = " + Page.User.IsInRole("reader"));
Response.Write("<li> Currently logged on user = " + Page.User.Identity.Name);
Response.Write("<li>admin = " + Page.User.IsInRole("admin"));
Response.Write("<li>reader = " + Page.User.IsInRole("reader"));
```

To make it simple, a base class page can be written for Admin, and deal with the authority in this page.

The function of customer service is used for Online communication between users, which is divided into web site client and customer service client. A real-time platform for communications between students, teachers, teaching director and teaching administrators should be provided to turn teaching and research into on the test database directly and maintain it dynamically.

## 4 Conclusion

With the dynamic running of the paper module, test module, communication module and access module in the system, the generating system and method of the paper makes the most of the positive effect of human resources , making accurate judgements on the studying state of students, and providing targeted papers to guide the implementation of learning and formative assessment.

### Acknowledgments

**Fund:** 2011, Shaoyang City, science and technology projects funded(G1119).

**Author:** Liu Yibing (1964 -), male, Hunan Shaodong, senior engineer, associate professor, Master. Research: electronic technology.

### References

1. Bo, Z., High, X.: Bank Management System Research and Implementation. Dalian University 27(2), 68–71 (2006)
2. Liu, Branch, Zhang, X.-Q., Zhang, pure gold: Intelligent systems in the test paper test paper Strategy Study. Computer and Information Technology 18(5), 40–42 (2006)

# Thermal Power Plant Pressure Dense Phase Pneumatic Ash PLC Control System Design and Implementation

Binlin Li<sup>1</sup>, Wen You<sup>1</sup>, Dianbin Song<sup>2</sup>, and Quan Tao<sup>1</sup>

<sup>1</sup> College of Electrical & Electronic Engineering, ChangChun University of Technology, Changchun, 130012, China

<sup>2</sup> Longhua Electric Power Yanji Electric Power Company Limited  
Yanji, 133000, China

{libinglin,youwen}@mail.ccut.edu.cn, 857823308@qq.com,  
tq666100@163.com

**Abstract.** Ash handling system of Coal-fired power plant is a advanced, economic, environmental science and technology. As the requirements and limitations of environmental protection, water resources and others, positive pressure pneumatic ash handling is widely used as ash handling method in China's coal-fired power plant. In this paper, Changchun High-tech thermal power plant unit 2 25Mw Ash delivery system is taken as object describing the configuration and project delivery process of pneumatic ash handling system, according to the control performance requirements of the ash handling system unit designed to using Siemens S7-200 type PLC for the control core of dual-machine redundant automatic control system to achieve the ash system control requirements and the main function, while in the PC ,software Wincc6.2 used to control PLC running to achieve a successful running of equipment data collection, centralized operations, real-time monitoring, data storage, fault alarm functions , improving the system's control and management capabilities. Reduced the labor intensity of running, saving the cost of maintenance.

**Keywords:** PLC, positive pressure dense phase, pneumatic ash, automatic control.

## 1 Introduction

Pneumatic ash handling is a kind of way taking air as the carrier, with pressure (positive or negative pressure) equipment and pipelines for transportation of powder material. Its technically is more advanced, reliable, low failure rate and less maintenance. In the environment, conservation of water resources and to achieve automatic control it has an incomparable advantage compared to traditional hydraulic ash delivery and conventional mechanical ash delivery way, Pneumatic ash handling has gradually become a trend and mandatory requirements in the power system. As the PLC internal logic control by microprocessor with high reliability, strong anti-interference, and other powerful feature-rich technology, it has been widely used in the production process to achieve automation control. This paper analyzes the design of pneumatic conveying and engineering configuration, process of pneumatic



conveying systems in power plant. According to the control requirements of Ash delivery system, using s7-200PLC and IPC to achieve dual- machine redundancy to achieve the development and design of a positive pressure dense phase pneumatic ash control system, the system has been successfully applied to pneumatic ash delivery control system in a power plant.

## 2 Scheme Design and Engineering Configuration

A power plant phase II has 2 x 25 MW boilers. Per boiler in the cloth bag dust collector is divided into A, B two side, each side has two ash hoppers. Dust collection of fly ash through the small by pipeline to ash pump loses ash library emissions. Precipitator fly ash collected by a small warehouse pump to Ash through ash discharge pipe.

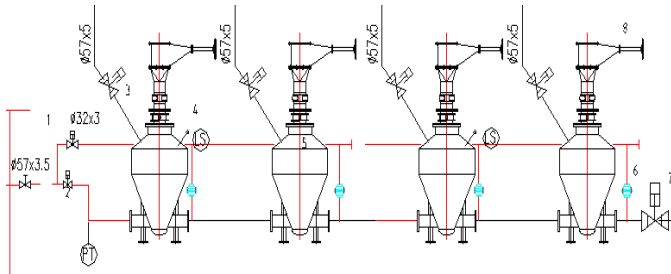
### 2.1 Conveyor Design

- Ash ways: thermal power plant way pneumatic ash conveying compressed air per kg ash weight ratio of the classification. Can be divided into three categories, namely, dilute phase, the phase and dense phase conveying. As the dilute phase and the phase transfer speed is high, the pipeline, valve wear, energy consumption is too high, so this project means the use of dense phase conveying.
- Sump pumps choices: Sump pumps (some said transmitter) is a pneumatic ash power equipment. Sump pumps are commonly used on the lead type, under the lead type, fluidization sump pumps. Under the lead-type sump pumps type structure is more simple on the lead, ash from the filter pump at the bottom of the hopper directly into warehouse delivery, without gray suspension operation, relying on air pressure to transport the ash, is the plunger type conveyor. The sump pumps feature a high concentration of dry ash conveying, air pressure can be almost entirely for transportation, transportation capacity. In addition, under the lead-type sump pumps internal structure is very simple, basically empty structures, without maintenance or little maintenance work. The project of ash conveying distance is short, choose the next lead-type sump pumps.
- Operation mode: Operation mode has intermittent transmission mode and continuous transmission mode. Intermittent transmission method according to arrangement of the sump pumps can be divided into modular, multi-system and intermediate positions such as operating mode. This project uses four sump pumps running in series. As the number of sump pumps running in series, if the storage volume of the pump is too large, will cause a large amount of ash, it is prone to plugging. To this end the project design storage volume of the pump is only 0.5 m<sup>3</sup>. [1]
- Valves: pneumatic ash system of valves used in a lot, which feed, the material commonly used pneumatic butterfly valve. As the pneumatic ash conveying system, the dry ash at high concentrations, high temperature, speed, coupled with these valves start frequently, so wear leak valve pneumatic ash conveying system is one of the most prominent problem. In order to solve these problems the project adopts the valve wear, dome valve. Dome valve is the biggest

characteristic of the valve actuating member and the static part are not hard contact, but on the sealing ring inflatable seal, and the sealing ring is provided with a cooling device, thus greatly improving the service life of the valve, and a valve even wear, only the sealing ring. [2]

- Blow block device: blow blocking device is also the pneumatic conveyance of the grey system a necessary equipment. Common methods currently blocking the main man blowing blown blocking, automatic blowing and suction three kinds of blocking. This project uses automatic blowing blocking mode. Automatic blocking device is blowing every 20 m along the gray tube at approximately a blow plugging, plugging and blowing compressed air blown master control block is used with a pressure switch solenoid valve connections. When the gray possession plugging tendency, its pressure increases, when the pressure rises to the set upper limit, the solenoid valve automatically opens to blow block. When the gray pipe flow recovery after its pressure is reduced, the electromagnetic valve closes automatically.

### 2.2 Delivery Pipeline and System Layout



1 -feed valve 2- intake valve 3- compensation valve

4-level meter 5- Dome discharging valve 6- Row plugging valve 7- discharge valve  
8 - Manual repair valve

Fig. 1. System layout diagram

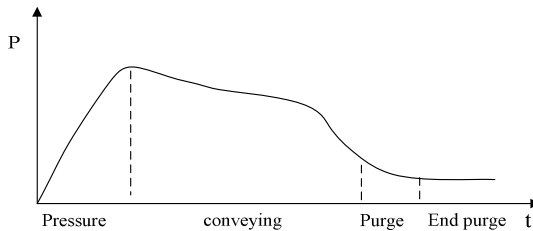
### 2.3 Pneumatic Ash System Engineering Configuration

Ash system in four sub-systems: PLC program control and SCADA systems, air compressors and air drying system, storage pump system, Ash system. Main configuration equipment: PLC program control cabinet 2 sets, 2 sets of running to operate the station, under storage bag filter pump 8 sets, air compressor equipment together with 1 2 3 sets, with a freeze dryer 2 of 3 sets of equipment, ash 2 sets of libraries and other appropriate equipment. PLC program control system for automatic run-based, supplemented by local manual, in the central control room to achieve centralized monitoring and management of PC. Sump pumps, air compressors and air drying systems are included in PC management. [3]

### 3 Positive Pressure Dense Phase Pneumatic Ash Conveying Systems Process Flow

Positive pressure dense phase pneumatic ash conveying system is based on solid-gas two-phase pneumatic conveying principle, the use of compressed air static pressure and dynamic pressure, high density, high efficiency transport of materials. This project uses sump pumps intermittent ash method, each conveying a pump fly ash as a work cycle, each work cycle consists of four sessions:

**1) feeding stages:** the feed stage, to determine the air pressure to the sump pumps the feed valve, balance valve is open, the intake valve and the valve is closed, dry ash from the filter into the storage hopper pump. When the pump ash rises above the material level or the feed reaches the set time, the system automatically close the inlet valve, balance valve, the feed end of the period.



**Fig. 2.** Typical curves of storage pump pressure

**2) Pressure stage:** open the discharge valve, open the place up after the delay 4 seconds to open the valve, delay 5 seconds to open the intake valve, the treated air supply clean compressed air through the flow control valve into the sump pumps, transfer person transmission phase.

**3) Conveying stages:** in the conveying stage, the discharge valve, the intake valve is open, the feed valve, balance valve is turned off, positions start pumping time 20 seconds. At this point the pump side of the inlet side of warehouse ash. After 20 seconds if you still meet the ash pressure, timing ash 300 seconds. In the ash during storage to maintain the basic stability of the pump pressure (see Figure 2). When the fly ash storage pump delivery is completed, the pipeline resistance decreased, decreasing the pressure inside the pipe. When sump pumps to lower the pressure within the pressure after that transfer phase is completed, enter the purge stage.

**4) purging phase:** when the sump pumps to lower the pressure within the pressure after that convey stages is completed, purge timer started. Purge phase feed valve, balance valve is still closed, fill valve, intake valve, the valve is open. When the purge

time is reached, program control valve first shut up, shut down after 5 seconds delay 10 seconds into the valve and close the valve. Major role in this stage is pure compressed air to give as gifts Ash residue of fly ash, the last was a pure air flow, system resistance decreased to a stable value. Air compressed air piping cleaning, cleaned up after, then enter the next cycle of work. [4]

### 4 Control System

The failure of the control system is likely to cause pressure dense-phase pneumatic ash system positions the pump plugging, pipeline plugging and other failures, thus requiring the ash system controller with fault tolerance. Common solution is to use high-end, such as S7-400 PLC to achieve, but the ash system requirements both in the control points, or in the control performance requirements, etc., generally without the use of high-end PLC. Based on the positive pressure dense phase pneumatic ash conveying system control requirements, the ordinary PLC and IPC to achieve double redundancy approach, redundant design of the control system shown in Figure 3. The system consists of PC, PLC control panels, field devices and other components.

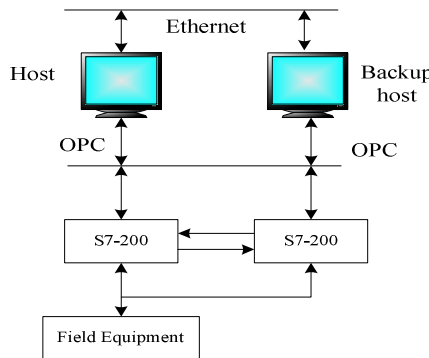


Fig. 3. System Topology

PC using two Advantech industrial control computer, running WINCC6.2 software to monitor the operation of PLC. Two industrial computers, a host, another for preparing machine, through TCP / IP network connection. Under normal circumstances the host by OPC monitor PLC and field operation of the device, producing the machine through the Ethernet data read from the host, in the query condition, while continuing to host the heartbeat signal detection; preparing machine once detected anomaly, will be in a very short period of time instead of the host, collected under the crew run data and control machine operation, the full realization of host control function, and continues to host a host when testing, repair, restoration of normal, making machine to immediately stop the next-bit machine communication, restored to query status. At the

same time the IPC by means of OPC and DCS networking factory, become a factory to a reliable system, in order to achieve power plant control integration requirements.

PLC control cabinet containing two sets of Siemens S7-200 CPU226, configuration of the same. Field device main gas source, sump pumps, a variety of auxiliary valves and sensors. All the input and output contacts and connected to the two PLC corresponding input output contact, i.e., an input contact at the same time to the two PLC input, two PLC and parallel to a live output contact output. Field output contacts are always open. Two PLC communication transmission of heartbeat signal. Two PLC communication addresses are the same, under normal circumstances the host PLC and host computer communication, to a PC to send all kinds of operation data and accept the superior machine control, and automatic control of the field equipment operation; preparation of PLC due to detection of the main PLC issued a heartbeat signal, receives the PC to send signal processing, the output contacts are in the release state, the field device is not controlled, in a state of readiness; when preparing PLC detection to the main PLC fault, it sends a signal to the main PLC, the main PLC to stop the communication with the host computer, the release of the output contacts, stop control field equipment, preparation of PLC over the main PLC functions, completely replace the main PLC. For the main, preparation of PLC operation set parameters can synchronously, main, preparation of PLC has its own device identification, computer detects the device identification name changes, send the running parameter set to the next machine, synchronous PLC operating parameters, to complete the preparation of PLC to the main PLC seamless transitions.

## 5 Software Design

The software includes PLC software and PC configuration software.

This system is a typical sequence control. Using Sequential Function Chart Method. Under the control requirements, the system's working process is divided into several stages, these are called "step." Step is divided according to the state of the output, the output control program in the M and Q in the program between the very simple logic. Application of sequential function chart (SFC) design method significantly reduces the design cycle, so that debugging and modifying is very easy, and improve design efficiency.

PC monitoring interface have pneumatic ash conveying system the main picture, parameter setting, data report screen picture, alarm display, user management, realize the main picture, standby redundant system, it can achieve the bin pump parameters setting, fault alarm.[5]

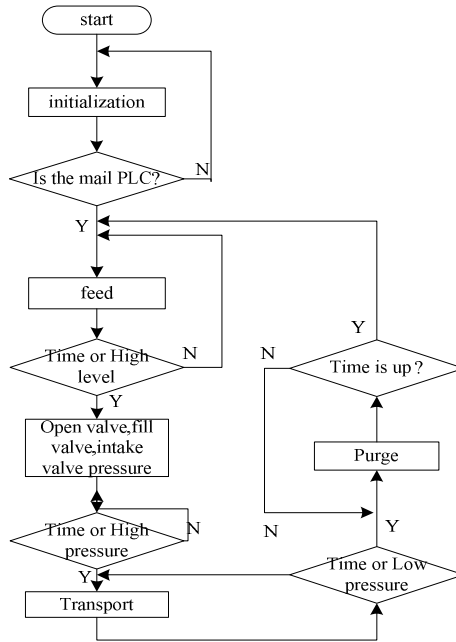


Fig. 4. Flow chart of software PLC

## 6 Conclusion

The pneumatic ash control system using a common PLC and IPC to achieve double redundancy of the methods to achieve positive pressure dense phase pneumatic ash conveying system. The pneumatic ash control system has been put into use in the Changchun one power plant, so far running well. System easy to operate, good stability, low gas consumption, wear and tear is small, with a high degree of automation, economic and practical features.

The system has running stable, full-featured, high reliability and many other features. Has some reference value for the same type Ash delivery system's optimization design, improve and enhance the reliability and economy of ash removal system.

## References

1. Fu, L., Ling, H.: The Technology of Pneumatic Ash in Thermal Power Plant. Guangxi Electric Power (3) (2003) (in Chinese)
2. Chen, C., Liu, B., Li, Z.: The application of Positive pressure dense phase pneumatic ash removal system technology in boiler industrial boiler (2) (2009) (in Chinese)

3. Wu, H., Liang, Y., Wu, J.: Multi-pump system pressure dense-phase pneumatic ash system PLC design *Micro-computer information* 24(19) (2008) (in Chinese)
4. Konovalov, Yashkin, Ermakov: Automated System For Removal And Pneumatic Transport Of Fly Ash From Electric Precipitator Hoppers *Power Technology and Engineering* 42(2) (2008)
5. Tang, R., Yao, Z., Eric, Y., Kun, M.: The Application of Siemens PLC in the bag dust removal system. *Automation Instrument* (7) (2005 ) (in Chinese)

# Communication Design for I2C Bus Based on Proteus

Fengjie Wu<sup>1</sup>, Liming Wu<sup>2</sup>, and Tangmei He<sup>1</sup>

<sup>1</sup> Laboratory Center of Guangzhou University, Guangzhou in China

<sup>2</sup> Information Engineering School of Guangdong University of Technology,  
Guangzhou in China

wufengjie501@163.com, {liming0118, 385357885}@qq.com

**Abstract.** We introduce Proteus virtual development techniques into the microcontroller I2C bus communication designs, and take the AT24C02 for example to explain the Proteus simulation software and hardware design circuit of I2C bus communication. Then we analyse the hardware and software problem in Proteus debugging and provide the solution. The Simulation results which is identical with theoretical analysis prove that the correctness and the accuracy of Proteus virtual simulation. It has very important significance in the microcontroller virtual development.

**Keywords:** I2C Bus, Microcontroller, Virtual Technology, Proteus.

## 1 Introduction

I2C serial bus is one of the bus standards that are widely used in the field of microelectronics communication control. It uses two transmission lines SDA and SCL to establish contact and transmission of information with microcontroller or peripheral devices. I2C serial bus is mainly used in microcontroller system expansion and multi-machine communication, with its simple and effectiveness advantages. [1]

Different from traditional hardware simulation, we introduce the microcontroller virtual development techniques in the design of the I2C bus communication, use the Proteus software which is strong in simulation and analysis to simulate the microcontroller and its peripheral equipment, and design and modify the hardware and software for multiple I2C devices with computer. The success in the application of technology forms a bran-new system design concepts, its design process is as follows: schematic design and simulation - circuit modification and improvement - to draw the printed circuit board - hardware assembly and debugging. It breaks the traditional system design mode, and makes the system design flaws exposed early in the initial stage of design. The design process overcomes the deficiency in continuously modifying the design of the hardware model.



## 2 About Proteus Simulation Software

Proteus is a circuit analysis and physical simulation software, which is running on the Windows operating system, is put out by a British company named Labcenter. It can simulate and analyze all kinds of analog circuits and integrated circuits. With its large advantages such as the rich experiment resources, the short experimental period, less investment in hardware, small loss in the experiment, and the proximity of the actual design. [2]

Proteus mainly includes the ISIS and the ARES. The main functions of ISIS are schematic design and simulation. And the ARES is mainly used in printed circuit board design. In addition to the features like other EDA tools, such as principles of layout, PCB automatic or manual routing, circuit simulation, the biggest feature of Proteus is the microcontroller-based design along with all the external circuit simulation. It can program on the virtual system's microcontroller directly, or use third-party application development and debugging environment in real-time virtual circuit debugging. It can predominate the operation of the program and the corresponding hardware circuit input and output states at any moment, when there is unconformity. It will track the Debug command in depth and find the problem and solution quickly.

In addition, Proteus also has the function of interactive circuit simulation. It can display system input and output by using dynamic model of peripherals such as keyboard, LED / LCD, etc., or combining with the virtual instruments of Proteus such as oscilloscopes, logic analyzers, etc.

## 3 Design Principles of I2C Bus Communication

### Hardware design

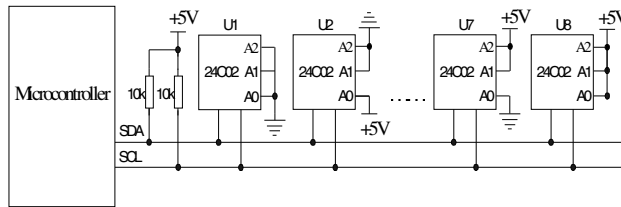
At present there are many types of devices that can be hung on the I2C bus. We take the AT24C02 for example. While designing, SDA and SCL of the devices are added to the same name access to the bus terminal respectively, and then connected with the SDA, SCL pins of the master. Then all devices connected to the I2C bus are controlled devices, while the bus master is the one that can control the controlled devices, such as microcontrollers, PLC, etc. And I2C bus allows multiple masters. [3]

The AT24C02 on the bus has a unique address  $1010xxx\mathbf{y}B$ , in which the high 4-bit code is a constant while being made in factories, xxx is based on level of pins A2, A1, A0, and y is the R / W signal. When y equals to 1 is read, otherwise is writing. Table 1 is the A2, A1, A0 encoding table. It shows that there are eight AT24C02 can be linked to the I2C bus at most. When designing, we can connect the VSS, TEST to the ground, and change the A2, A1, A0 level status according to Table 1.

**Table 1.** The AT24C02 address allocation table

NO.	A2	A1	A0	R/Y	R/W address
U1	1	0	0	0/1	0A0H/0A1H
U2	1	0	1	0/1	0A2H/0A3H
U3	1	0	1	0/1	0A4H/0A5H
U4	1	0	1	0/1	0A6H/0A7H
U5	1	0	1	0/1	0A8H/0A9H
U6	1	0	1	0/1	0AAH/0ABH
U7	1	0	1	0/1	0ACH/0ADH
U8	1	0	1	0/1	0AEH/0AFH

With comparison, we select AVR ATmega8 MCU as the master, which provides a fully compatible TWI I2C serial bus interface. Pins PC5/ADC5/SCL, PC4/ADC4/SDA are the clock signal and data signal of I2C bus respectively. Because they are open-drain pins, the circuit design requires an external pull-up resistor. The design block diagram is shown in Figure 1.

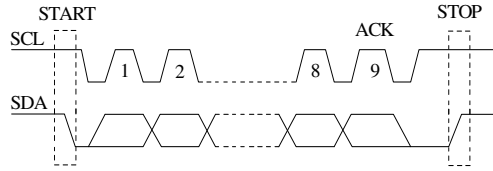


**Fig. 1.** Design schematics of hardware circuit

**Software design**

On the I2C bus, the master recognizes the controller's address through the use of software addressing. It calculates each I2C device's read and writing address according to the actual hardware circuit diagram. For example Table 1 is an address table corresponding to the I2C devices shown in Figure 1. In the process of code execution, the master explains the instructions contained the address, and read data from or writes data to the corresponding I2C devices. [4]

Figure 2 shows the timing diagram of I2C communication. The process can be discussed in three steps.



**Fig. 2.** I2C bus data transfer sequence

(1) Start signal: sent by the master, that is SDA changes from high level to low level while SCL is high;

(2)The address transmission or data transmission: First, address frame of I2C device sent by the master is made up of 7-bit address signal and a 1-bit reading-and-writing. When the read and write signal is 1 it is reading, otherwise is writing. The I2C device on the bus with the same address sent by the master is selected and returns a response signal to the master. Then, after receiving the response signal, the master starts to read data from or write data to the I2C device. The data transmission which is the same as the address transmission transfers in bytes and the length of each byte is 8. The first transmission is MSB while LSB the last is. The transmission of each byte is followed by a response signal, when the response signal is Ack, it means continuing to transmit data while Nack means the end of the case.

(3) Stop signal: sent by the master, that is SDA changes from low level to high level while SCL is high;

The system software is designed in modularization. Initialization program of the system, the read and write data program of I2C devices, and input or output device program that system may include are packaged as a separate function modules. All function modules are integrated to a complete system after each function module is successfully debugged, thereby reducing the difficulty of debugging the system while improving development efficiency. Source code of the program is written with a common build environment for AVR microcontroller ---- BASCOM-AVR, which contains a wealth of data types, a large number of simple and practical functions, and also develops some special peripheral functions, such as I2C statement, LCD statement, RC5 statements, etc.

Now, we take AVR microcontroller communicating with I2C device U8 for example, to show I2C bus operation subroutine of read and writing. From Table 1, it is clear that the device’s read address is & Haf and write address is & Hae.

```

1. Subroutine about microcontroller reads a data from I2C-bus
I2CStart
I2CWbyte &Haf
I2CWbyte 0
I2CStart
I2CWbyte &Haf
I2CRbyte X , Nack
I2CStop
    
```

```

2. Subroutine about microcontroller writes a data to I2C-bus
I2CStart
I2CWbyte &Hae
I2CWbyte 0
I2CWbyte X
I2CStop
Waitms 20
    
```

### 4 Circuit Simulation and Results Analysis

In order to achieve the interactive simulation, the design also includes some input and output devices. We will use the 8 × 1 keyboard instead of command unit in microcontroller control process to achieve the read and write operations of the I2C device U1 ~ U8, and use the LM1602 LCD to export the simulation results.

The introduction of the keyboard and LCD makes us finish a series of interactive experiments about the AT24C02 on the I2C bus. We take the data communication in a microcontroller control process for example. It requires pressing the button followed by K1 ~ K8 to write data  $10 \times n + m$  ( $0 \leq n \leq 7, 0 \leq m \leq 9$ ) to the AT24C02 U1 ~ U8 respectively, then read out the 10 newly written data after an appropriate delay, and display the read results by LCD U10.

Combine Figure 1 with the keyboard and display circuit, we use Proteus to design the simulation circuit which is shown in Figure 3.

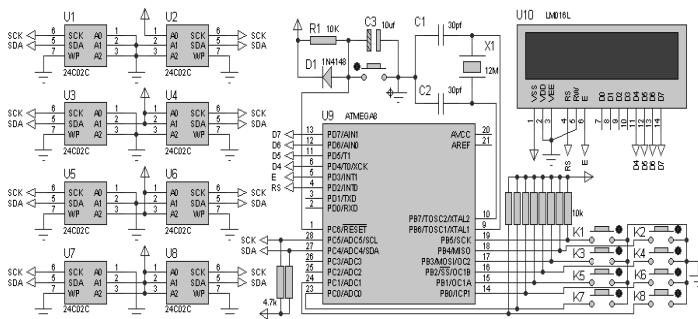


Fig. 3. Simulation circuit of I2C bus communication

Create the design program files by running BASCOM-AVR software. On the one hand, set the Options-related attributes in the output column by selecting HEX file; on the other hand, put the source code in the file and save it as \*.bas file, then compile the program and debug it, finally generate the \*.hex file. [5] After checking the correctness of the simulation circuit in Figure 3, load the generated \*.hex file on the microcontroller U9, and then run the Proteus I2C bus communication circuit in order to achieve a virtual simulation.

Operate buttons K1 ~ K8 respectively, the corresponding output results is shown on the LCD screen as in Table 2. It is obvious that the results are fully matching to the design requirements. And so is strongly proved the design correctness and feasibility

of the I2C bus communication by using Proteus. In addition, when pressing the button K8, the circuit simulation results are shown in Figure 4 screenshots.

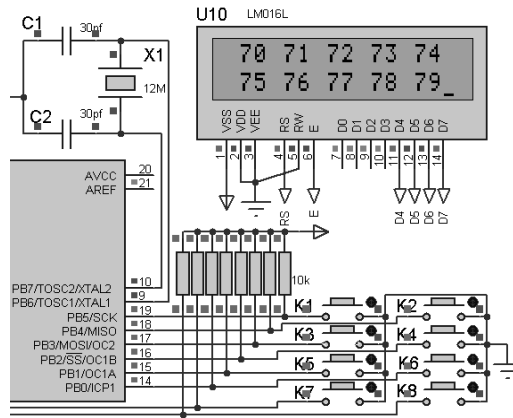


Fig. 4. Proteus simulation results screenshot

## 5 Debugging Common Problems and Solutions

As the user is not thoroughly familiar with the I2C bus works, or doesn't operate properly to the Proteus simulation software or other reasons, there will be a range of hardware or software problems in the design process. We list some common points as follows, and solve them step by step according to the failure resulting from primary and secondary factors.

### 1) No response to the circuit simulation when operating all the keys

Firstly, check whether the microcontroller U9 has been loaded the \*.hex program file correctly. Secondly, operate buttons K1 ~ K8 successively and exam the level change of each key. The correct change should be high level - low level - high level, that is, it is the level stage of the low level only when the button is pressed, otherwise is high level. Thirdly, examine pins A2, A1, A0 of I2C devices U1 ~ U8 respectively, the correct level should be consistent with table 1. And at the same time, exam SDA and SCL level is whether the same with microcontroller's. Fourthly, check whether the LCD screen is working properly by loading a proper LCD display program to microcontroller U9. If it can correctly display, the LCD screen has no failure. Finally, check the design's program source code whether it is correct by using debug commands to single-step debug or debug trace.

### 2) No response to the circuit simulation when operating individual key

Take no response to simulate circuit when operating the key K1 for example. It can, above all, rule out the failure possibilities of microcontroller U9 and LCD screen U10. Firstly, operate button K1 and examine its level change. The correct change should be high level - low level - high level, that is, it is the level stage of the low level only when the button is pressed, otherwise is high level. Secondly, examine pins

A2, A1, A0 of I2C devices U1 respectively, the correct level should be consistent with table 1. And at the same time, exam SDA and SCL level is whether the same with microcontroller's.

### 3) Error results in the circuit simulation after key operation

Firstly, check whether the LCD screen is working properly by loading a proper LCD display program to microcontroller U9. If it can correctly display, show the LCD screen has no failure. Secondly, check the design of the program source code, and give emphasis to mainly examine the program of liquid crystal display and read and write operating of I2C device.

## 6 Conclusion

We describe the design of AVR microcontroller I2C bus communication module. It contains hardware design and realization of Proteus simulation circuit, uses BASCOM-AVR software to complete Software Programming, and has solved some common hardware and software problems. Simulation results show that the use of EDA tools Proteus for microcontroller virtual system design has many advantages, such as high efficiency, low cost, large proximity of the actual design, etc., and has great significance for microcontroller design and development. This method is particularly suitable for beginners to use SCM, and is also a powerful complement of SCM courses in teaching design experiment.

## References

1. Chen, G.-J., Jia, J.-L.: Design of I2C bus system based on SCM. *Chinese Journal of Scientific Instrument*, 2465–2466, 2472 (June 2006)
2. Li, N., Liu, Y.-J.: Application of Proteus on Simulation for MCU. *Modern Electronic Technique*, 181–182 (April 2007)
3. Zhang, L., Liao, J.-B., Kong, L.-W.: Simulation of SCM I2C bus communication with the digital potentiometer using Proteus. *Electronic Measurement Technology*, 49–52 (November 2010)
4. Chen, Z.-H.: The realization of FC-Bus to MCS51 Single Chip Microcontroller family data gather system. *Control & Automation*, 67–68 (January 2005)
5. Wu, F.-J., Xie, C.-Y., Huang, W.-K.: AVR Microcontroller virtual system design based on Proteus. *Computer Knowledge and Technology*, 1066–1067 (August 2008)

# Application of Proteus in Microcontroller Comprehensive Design Projects

Fengjie Wu and Tangmei He

Laboratory Center of Guangzhou University, Guangzhou in China  
wufengjie501@163.com, 385357885@qq.com

**Abstract.** We introduce the simulation method of the Proteus software, and propose the virtual system design which based on the Proteus. Combining with the example of the smallest microcontroller system, we describe the specific application methods of the Proteus in experimental teaching. And finally, we achieve the soft parallel hardware co-development of the microcontroller comprehensive design projects, which can effectively save hardware resources and improve the development efficiency. It has very important significance in the microcontroller experimental teaching.

**Keywords:** Proteus, Virtual Simulation, Microcontroller, Experimental Teaching.

## 1 Introduction

In the traditional experimental teaching, a system design model in the microcontroller comprehensive design projects is: schematic design - physical circuit welding - hardware debugs and modify. Because of the limitations in ability and knowledge of the students, a range of hardware problems, such as short circuit and open circuit, pad out, installation errors or damage to other components, often appear as the result of modifying the hardware circuit for repeatedly. They, thereby, affect the system stability and reliability. It is obviously difficult to achieve the desired effect of teaching.

Use Proteus software with its strong function in the simulation and analysis to simulate the microcontroller and its peripheral equipment. Design, debug and modify hardware and software of the microcontroller system on the computer. Draw the physical circuit after the successful circuit simulation. [1] It forms a new system design model is: schematic design - simulation circuit design and modification - physical production and testing. The model makes the flaws exposed early in the design, shorts design cycles and reduces experimental costs. While also maximizing student's learning initiative and innovation. It is of great significance in developing innovative and practical talent, reforming the traditional teaching model and improving the quality of experimental teaching.

## **2 About Proteus Simulation Software**

Proteus is a circuit analysis and physical simulation software launched by British Lab center Company, which run on Windows platform and is made up mainly by the ISIS and ARES. Isis's main function is to schematic design and simulation, while ARES is mainly used for printed circuit board design. [2] Its main features are as follows.

### **2.1 Friendly User-Interface**

Proteus main interface contains the system menu command bar, library resources, parts list bar, drawing workspace, etc. The use of their operating commands is the same as the other software of the Windows. Simulation of the components and test instruments is very close to the physical appearance, and the operating is in much the same. The user-interface is friendly and intuitive to learn.

### **2.2 Wealth of Experimental Resources**

Proteus component library contains a large number of components, and they are stored in 36 different categories in the library. It includes basic components, semiconductors, microcontroller components, COMS and TTL components, PLD and FPGA components, debugging tools, storage devices, integrated operational amplifiers, ADC and DAC devices, etc. The user can expand the existing components library through the product line supplier, and also can modify or create libraries according to the design needs.

### **2.3 Powerful Virtual Instrument**

Proteus provides 12 kinds of debugging and testing instruments commonly used in the model, such as oscilloscope, logic analyzer, frequency meter, SPI and I2C debugger, generator, AC and DC voltage and ammeter, etc. These virtual instruments not only are very similar with the physical in appearance, but also almost the same in use. And they are easily used in virtual testing and virtual system debugging.

### **2.4 Unique Simulation Approach**

In addition to schematic layout, PCB layout, circuit simulation and other features, the biggest feature of Proteus is able to simulate the microcontroller and its peripheral devices[3], and program directly on the MCU in circuit simulation or introduce the third-party application developers and debugging environment to debug hardware and software of the virtual circuit. Proteus also has the function of interactive circuit simulation and can control the input and output status in real-time by a combination of the dynamic model and the virtual instruments of the peripheral software.

## **3 System Design Method Based on Proteus**

Proteus virtual application development technology for the microcontroller comprehensive design projects includes five steps: the development of system solutions, Proteus circuit design, virtual system simulation, physical production and testing.



Development of system solutions: the student collects, collates and summarizes the project information according to the synopsis of the project design mission statement. Design the development of overall system preliminarily according to the preliminary design requirements of the design, including microcontroller models, software development platform, choice of programming languages, and the overall block diagram of the hardware and software flow chart of the total formulation, etc. Demonstrate the program and turn over to the instructor for review.

Proteus circuit design: it includes simulation circuit design and software programming. On the one hand, design the virtual simulation circuit with the modular design method which generally includes the microcontroller core module, the basic input-output modules, each functional sub-module, etc. Design and draw the hardware schematic with Proteus, at the same time, detect the sub-module circuit accuracy. On the other hand, design software program according to the functions of each module, make the main program of the system and the module subroutine flow chart. Program and debug source code in the Proteus programming environment or a third-party application development tools, such as Keil, BASCOM-AVR, etc.

Virtual System Simulation: Each functional modules runs software and hardware co-simulation with the core modules of the MCU respectively after the building of the system's hardware model and software model. And it achieves through loading the \*.hex file to the MCU. Integrate into a complete microcontroller system after the successful simulation of the sub-module. Use LED, LCD or other basic output device to display the simulation results, and use the virtual instruments configured in Proteus, such as oscilloscopes, I2C debugger, voltmeter, .etc, to measure the input and output status of the circuit components, or the changes in the signal circuit nodes.

Physical production and testing: Produce PCB and install related components by welding hardware circuit according to the Proteus simulation circuit diagram or using network relations of ISIS and ARES to generate PCB automatically in ARES. Record the output of the hardware circuit, use the physical measurement and instrumentation to test the nodes and the components corresponding to Proteus, and then, compare the physical test with the simulation results, if there is deviation, investigate the physical circuit first of all, such as whether the line is correct, the components are damaged or the parameters are correct, etc., then check the correctness and the feasibility of the Proteus circuit simulation.

## **4 Proteus Simulation Circuit Design Example**

Take the smallest microcontroller system for example to describe the design process of the microcontroller comprehensive design projects with Proteus.

### **4.1 Hardware Design**

Smallest microcontroller system is a preferred comprehensive design projects in the microcontroller courses, including MCU and the basic input and output devices, and there may be other peripherals. For example, the smallest microcontroller system described in the paper includes microcontroller, keyboard, LCD, I2C-bus devices and data collection devices, etc., mainly to complete the data acquisition, displaying and

storage through the keyboard controls microcontroller. The block diagram of the system is shown in Figure 1.

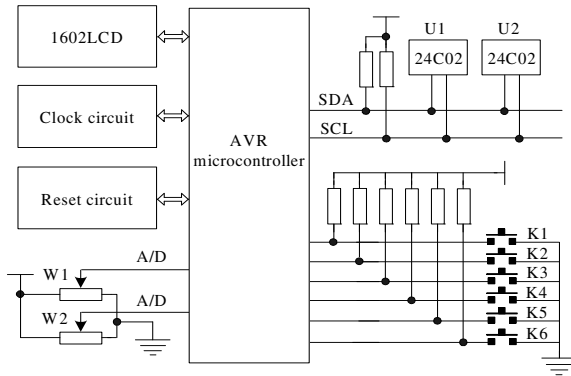





Fig. 1. Smallest microcontroller system block diagram

According to Figure 1, use Proteus to design the circuit diagram of each functional module through the following five steps. Firstly, click the button  to enter the main interface of Proteus and new design, set the basic properties of the design and save it as \*. DSN files. Secondly, click the button  to enter the components search window to find all the devices for the design, and add to the components list column. Thirdly, add the various components required for the functional module to the drawing area, distribute rationally and set the parameters. Fourthly, click the button  to connect, when the mouse pointer nears the pin is showing 'x', left-click to draw lines, but right-click to cancel the operation. Finally, check the correctness of the circuit for each functional module.

Based on the above steps, the simulation circuit diagram of Proteus is shown in Figure 2.

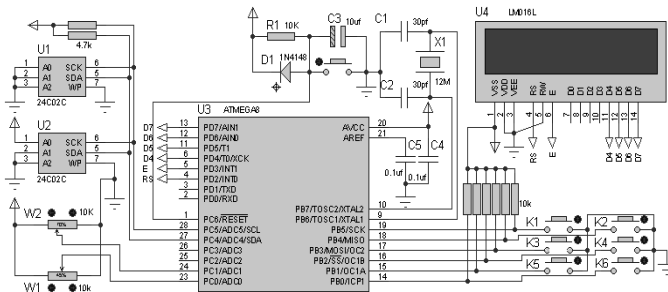


Fig. 2. Proteus simulation circuit of the smallest microcontroller system

## 4.2 Software Design

It is very simple in the hardware circuit simulation in Proteus, as long as load \*.hex files in the microcontroller and click the run button in Proteus main interface.



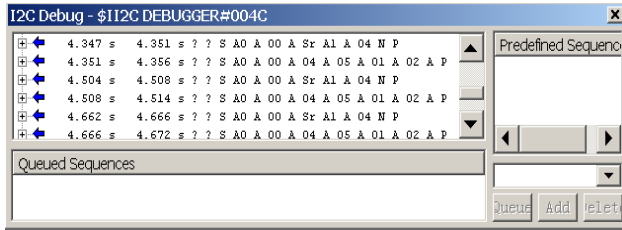


Fig. 4. The debug result of I2Cdebugger

According to microcontroller measurement requirements, generally need to scale conversion of the measure results. After the A / D sampling, so it is often necessary to convert the value to voltage values. As the AVR ATmega8 built a 10-bit A / D converter, then calculate the analog input signal voltage through the formula (1) calculated. [5]

$$V_i = V_{ref} \times X / (2^{10} - 1) \tag{1}$$

Among them,  $V_i$  is the analog input signal voltage,  $V_{ref}$  is the reference voltage for the ADC conversion, and  $X$  is the ADC conversion value.

Adjust potentiometer W1, and record the ADC results respectively. Calculate the voltage value of the analog input signal according to the formula (1), and use voltage probe W1(3) to real-time measure the voltage value of the corresponding analog input signal, among the three relationship shown in Table 1.

Table 1. Smallest microcontroller system simulation results in table

ADC conversion result	102	205	307	410	512
Input voltage calculated	0.4985	1.0020	1.5005	2.0039	2.5024
Input voltage measured	0.5	1.0	1.5	2.0	2.5
ADC conversion result	614	717	819	922	1023
Input voltage calculated	3.0010	3.5044	4.0029	4.5064	5.0000
Input voltage measured	3.0	3.5	4.0	4.5	5.0

From Table 1, verifiable AVR ATmega8 has a high accuracy built-in ADC, its absolute accuracy is  $\pm 2LSB$  and nonlinearity is  $0.5LSB$ . Compare simulation results with theoretical analysis, the consistency between them demonstrates the correctness and feasibility of the use of Proteus for the microcontroller comprehensive design projects.

## 6 Conclusion

The use of EDA tools Proteus for experiment teaching of the microcontroller comprehensive design projects has changed the traditional design experiment teaching, which relies on material means, such as electronic components and instrumentation equipment, etc. It can effectively short the test cycle, reduce the

hardware investment, decrease the loss, and get close to the actual design. So Proteus is an important learning and development tools for microcontroller technology. Practice shows that the introduction of the virtual development technology in the microcontroller course will help the students mobilize the enthusiasm of independent learning and initiative, significantly improve their circuit design and system development capabilities, and cultivate their sense of innovation and the spirit of rigorous scientific research develop.

## References

1. Wang, H.-C., Zhang, J.-J.: Proteus-based microcontroller multi-channel temperature acquisition system simulation design. *Journal of Shaanxi University of Science Technology* (2007)
2. Peng, W.: *Microcontroller C Programming Language Training 100 cases — AVR + Proteus based simulation*, pp. 35–44. Beijing Aerospace University Press (2010)
3. Li, L.: Simulation Design of Data Acquisition System Based on Proteus. *Journal of Electric Power* (6), 226–227 (2009)
4. Liu, X.-Y.: Application of I2C Debugger in MCU Designing Examination. *Journal of Gannan Normal University* (6), 105–108 (2009)
5. Zhou, J.-X., Ma, C., Geng, D.-G.: *AVR microcontroller BASIC language programming and development*, pp. 146–147. Tsinghua University Press surgery (2005)

# Modelling Three-Dimensional Foundation Pit Using Python Scripts

Shi-lun Feng<sup>1</sup>, Yong Han<sup>2</sup>, Pu-lin Li<sup>3</sup>, and Jun Li<sup>3</sup>

<sup>1</sup> Institute of Civil Engineering, Tianjin University, Tianjin 300072, China

<sup>2</sup> China North Municipal Engineering Design and Research Institute, Tianjin 300074, China

<sup>3</sup> Tianjin Institute of Geotechnical Investigation and Surveying, Tianjin 300191, China

shilunfeng@126.com, {452160284, 190462173}@qq.com,

lijun411198@163.com

**Abstract.** Accompanying with rapid development of city construction, there are more and deeper foundation pit projects in city centre, structure calculation of the foundation pit is of important significance. This paper uses a Python script to produce an input file which is compatible with the ABAQUS in order to make the foundation pit model easier and faster. And an engineering example is presented to prove the method feasible.

**Keywords:** Python, foundation pit, ABAQUS.

## 1 Introduction

Accompanying with rapid development of city construction, there are more and deeper foundation pit projects in city centre, structure calculation of the foundation pit is of important significance. Due to the complexity of the foundation pit and the surrounding buildings, numerical methods are often used to evaluate the safety of foundation pit and the surrounding buildings.

Yunian Lu et al[1] have studied the behavior and performance of retaining structures and bracing systems using Abaqus. Haizhu Guo et al[2] have analyzed the influence of supporting time on deformation of retaining structure and internal force by Abaqus. Jishu Sun et al[3] have established the finite element analysis model using PLAXIS software and put forward the influence characteristics and the laws of soil water content on the retaining structure bearing. Based on the deep foundation pit in a station on Suzhou Metro Line 2, Rulin Zhang et al[4] have carried out the numerical simulation by means of elastoplastic finite element program PLAXIS for the process of excavation and supporting. The factors of design, construction and natural environment on effects of the deformation of timbering structures were studied, including the depth of excavation, the rigidity of timbering structures, the rigidity of braces and parameters of soil mass etc. Yingyu Jin et al[5] have carried out mechanical analysis for the pit excavation and support using plastic-elastic hardening model to simulate constitutive relation of soil characteristics. Hongwei Ying et al[6] have proposed a method to analyze behaviors of beam-slab braced foundation pits considering the effect of construction process. A 3D FEM program based on the model has been developed.

ABAQUS is the most advanced nonlinear finite element analysis software and is suitable for foundation pit analysis[7]. In ABAQUS, establishment and analysis of models are mainly divided into several parts: parts, assembly, property, mesh, interaction, step, submit and so on. When a complex foundation pit is analyzed, many time and efforts need be consumed to make the model. In the present work, a Python script[8] is presented to produce an input file which is compatible with the ABAQUS in order to make the foundation pit model easier and faster.

## 2 Creating Three-Dimensional Foundation Pit Model by Python

We will use an engineering example to introduce the method of automatic modeling by Python. The engineering example is the JieFang Bridge foundation pit in Tianjin in China. The excavation depth is 12.9m and geometric size is 50.5m×35m. Thickness of the retaining wall is 0.8m, and the depth of the retaining wall is 24m. Parameters of soil layers are shown in table 1.

## 3 Creating Soil Layers

Because the soil is not a homogeneous soil layer, the soil must be modelled by various layers. So the soil layer must be created according to its thickness, which is a very troublesome work if it is done by people step by step. But it is very easy and fast for Python to do it. The following code excerpt shows how the various soil layers is created:

**Table 1.** Main physical and mechanical parameters of soil

Soil layer	$\gamma$ kN/m <sup>3</sup>	E MPa	c kPa	$\Phi$ °
miscellaneous fill	19.1	5.78	27.6	18.5
Silty	20.5	9.84	18.1	27.4
Silty clay	20.2	7.07	30.1	21.5
Silty	20.4	18.33	9.8	34.1
Silty clay	20.2	7.38	31.5	19.5
Silty	20.7	7.97	17.3	36.5

```

i=0
while i<len(q):
    print i
    s1.Line(point1=(-b/2, h/2-q[i]), point2=(b/2, h/2-q[i]))
    s1.HorizontalConstraint(entity=g[6+i])
    i=i+1
p = mdb.models['Model-1'].parts['Part-1']
f = p.faces

```

```

pickedFaces = f.getSequenceFromMask(mask=('[#10 ]', ), )
e1, d2 = p.edges, p.datums
ee1=e1.findAt(((b,h/2,a), ))
p.PartitionFaceBySketch(sketchUpEdge=ee1[0], faces=pickedFaces, sketch=s1)
s1.unsetPrimaryObject()
del mdb.models['Model-1'].sketches['__profile__']
i=0
while i<len(q):
    p = mdb.models['Model-1'].parts['Part-1']
    c = p.cells
    pickedCells = c.findAt(((0,0,0), ))
    e, d = p.edges, p.datums
    pickedEdges = e.findAt(((b/2,h-q[i],a), ))
    e9=e.findAt(((b,h,a/2), ))
    p.PartitionCellByExtrudeEdge(line=e9[0], cells=pickedCells, edges=pickedEdges,
sense=REVERSE)
    i=i+1
    continue
p = mdb.models['Model-1'].parts['Part-1']
f, e, d = p.faces, p.edges, p.datums
ff=f.findAt(((1,h,1), ))
ee=e.findAt(((b,h,a/2), ))
t = p.MakeSketchTransform(sketchPlane=ff[0], sketchUpEdge=ee[0],
    sketchPlaneSide=SIDE1, origin=(b/2, h, a/2))
s = mdb.models['Model-1'].ConstrainedSketch(name='__profile__',
    sheetSize=197.98, gridSpacing=4.94, transform=t)
g, v, d1, c = s.geometry, s.vertices, s.dimensions, s.constraints
s.setPrimaryObject(option=SUPERIMPOSE)
p = mdb.models['Model-1'].parts['Part-1']
p.projectReferencesOntoSketch(sketch=s, filter=COPLANAR_EDGES)
s.Line(point1=(-b/2, bb/2+w), point2=(b/2, bb/2+w))
s.HorizontalConstraint(entity=g[6])
s.Line(point1=(-b/2, bb/2), point2=(b/2, bb/2))
s.HorizontalConstraint(entity=g[7])
s.Line(point1=(-b/2, -bb/2), point2=(b/2, -bb/2))
s.HorizontalConstraint(entity=g[8])
s.Line(point1=(-b/2, -bb/2-w), point2=(b/2, -bb/2-w))
s.HorizontalConstraint(entity=g[9])
s.Line(point1=(-aa/2-w, a/2), point2=(-aa/2-w, -a/2))
s.VerticalConstraint(entity=g[10])
s.Line(point1=(-aa/2, a/2), point2=(-aa/2, -a/2))
s.VerticalConstraint(entity=g[11])
s.Line(point1=(aa/2, a/2), point2=(aa/2, -a/2))
s.VerticalConstraint(entity=g[12])
s.Line(point1=(aa/2+w, a/2), point2=(aa/2+w, -a/2))
s.VerticalConstraint(entity=g[13])
p = mdb.models['Model-1'].parts['Part-1']
f = p.faces
pickedFaces = f.getSequenceFromMask(mask=('[#4000000 ]', ), )
e2, d2 = p.edges, p.datums
ee2=e2.findAt(((b,h,a/5), ))
p.PartitionFaceBySketch(sketchUpEdge=ee2[0], faces=pickedFaces, sketch=s)

```



```

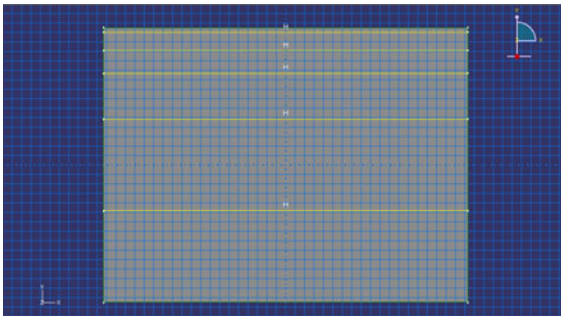
s.unsetPrimaryObject()
del mdb.models['Model-1'].sketches['__profile__']
p = mdb.models['Model-1'].parts['Part-1']
c = p.cells
pickedCells = c.getSequenceFromMask(mask=('[#7fffff ]', ), )
e, d = p.edges, p.datums
ew0=e.findAt(((b-aa-w)/2,h,0,))
ew1=e.findAt(((b/2-aa/2-w,h,a/4-bb/4-w/2,))
ew2=e.findAt(((b/2-aa/2-w,h,a/2-bb/2-w/2,))
ew3=e.findAt(((b/2-aa/2-w,h,a/2,))
ew4=e.findAt(((b/2-aa/2-w,h,(a+bb+w)/2, ))
ew5=e.findAt(((b/2-aa/2-w,h,3*a/4+bb/4+w/2,))
ew6=e.findAt(((b/2-aa/2-w/2,h,a,))
ew7=e.findAt(((b/2-aa/2,h,3*a/4+bb/4+w/2,))
ew8=e.findAt(((b/2-aa/2,h,(a+bb+w)/2,))
ew9=e.findAt(((b/2-aa/2,h,a/2,))
ew10=e.findAt(((b/2-aa/2,h,a/2-bb/2-w/2,))
ew11=e.findAt(((b/2-aa/2,h,a/4-bb/4-w/2,))
ew12=e.findAt(((0,2,0, ))
pickedEdges
=(ew0[0],ew1[0],ew2[0],ew3[0],ew4[0],ew5[0],ew6[0],ew7[0],ew8[0],ew9[0],ew10[0],ew11[0
])
p.PartitionCellByExtrudeEdge(line=ew12[0], cells=pickedCells, edges=pickedEdges,
sense=REVERSE)
p = mdb.models['Model-1'].parts['Part-1']
c = p.cells
pickedCells=c.getSequenceFromMask(mask=('[#ffffff:2 #1f ]', ), )
e1, d1 = p.edges, p.datums
ex0=e1.findAt(((b+aa+w)/2,h,0,))
ex1=e1.findAt(((b/2+aa/2,h,a/4-bb/4-w/2,))
ex2=e1.findAt(((b/2+aa/2,h,a/2-bb/2-w/2,))
ex3=e1.findAt(((b/2+aa/2,h,a/2,))
ex4=e1.findAt(((b/2+aa/2,h,(a+bb+w)/2, ))
ex5=e1.findAt(((b/2+aa/2,h,3*a/4+bb/4+w/2,))
ex6=e1.findAt(((b/2+aa/2+w/2,h,a,))
ex7=e1.findAt(((b/2+aa/2+w,h,3*a/4+bb/4+w/2,))
ex8=e1.findAt(((b/2+aa/2+w,h,(a+bb+w)/2,))
ex9=e1.findAt(((b/2+aa/2+w,h,a/2,))
ex10=e1.findAt(((b/2+aa/2+w,h,a/2-bb/2-w/2,))
ex11=e1.findAt(((b/2+aa/2+w,h,a/4-bb/4-w/2,))
ex12=e1.findAt(((0,2,0,))
pickedEdges =(ex0[0], ex1[0], ex2[0], ex3[0], ex4[0], ex5[0], ex6[0],
ex7[0], ex8[0], ex9[0], ex10[0], ex11[0])
p.PartitionCellByExtrudeEdge(line=ex12[0], cells=pickedCells, edges=pickedEdges,
sense=REVERSE)
p = mdb.models['Model-1'].parts['Part-1']
c = p.cells
pickedCells=c.getSequenceFromMask(mask=('[#ffffff:3 #7ffff ]', ), )
e2, d = p.edges, p.datums
ey0=e2.findAt(((0,h,a/2-bb/2-w/2,))
ey1=e2.findAt(((b/4+aa/4-w/2,h,a/2-bb/2,))
ey2=e2.findAt(((b/2-aa/2-w/2,h,a/2-bb/2,))

```

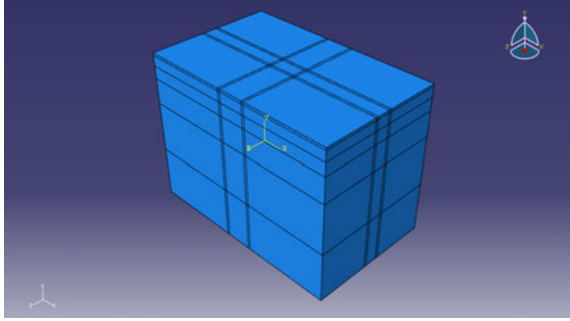
```

ey3=e2.findAt(((b/2,h,a/2-bb/2),))
ey4=e2.findAt(((b/2+aa/2+w/2,h,(a-bb)/2), ))
ey5=e2.findAt(((3*b/4+aa/4+w/2,h,a/2-bb/2),))
ey6=e2.findAt(((b,h,(a-bb-w)/2),))
ey7=e2.findAt(((3*b/4+aa/4+w/2,h,a/2-bb/2-w),))
ey8=e2.findAt(((b/2+aa/2+w,h,a/2-bb/2-w),))
ey9=e2.findAt(((b/2,h,a/2-bb/2-w),))
ey10=e2.findAt(((b/2-aa/2-w/2,h,a/2-bb/2-w),))
ey11=e2.findAt(((b/4+aa/4-w/2,h,a/2-bb/2-w),))
ey12=e2.findAt(((0,2,0),))
pickedEdges =(ey0[0], ey1[0], ey2[0], ey3[0], ey4[0], ey5[0], ey6[0], ey7[0],
ey8[0], ey9[0], ey10[0], ey11[0])
p.PartitionCellByExtrudeEdge(line=ey12[0], cells=pickedCells, edges=pickedEdges,
sense=REVERSE)
p = mdb.models['Model-1'].parts['Part-1']
c = p.cells
pickedCells=c.getSequenceFromMask(mask='[#ffffff:10 #1ffffff ]', )
e3, d1 = p.edges, p.datums
ez0=e3.findAt(((0,h,a/2+bb/2+w/2),))
ez1=e3.findAt(((b/4+aa/4-w/2,h,a/2+bb/2+w),))
ez2=e3.findAt(((b/2-aa/2-w/2,h,a/2+bb/2+w),))
ez3=e3.findAt(((b/2,h,a/2+bb/2+w),))
ez4=e3.findAt(((b/2+aa/2+w/2,h,(a+bb)/2+w), ))
ez5=e3.findAt(((3*b/4+aa/4+w/2,h,a/2+bb/2+w),))
ez6=e3.findAt(((b,h,(a+bb+w)/2),))
ez7=e3.findAt(((3*b/4+aa/4+w/2,h,a/2+bb/2),))
ez8=e3.findAt(((b/2+aa/2+w,h,a/2+bb/2),))
ez9=e3.findAt(((b/2,h,a/2+bb/2),))
ez10=e3.findAt(((b/2-aa/2-w/2,h,a/2+bb/2),))
ez11=e3.findAt(((b/4+aa/4-w/2,h,a/2+bb/2),))
ez12=e3.findAt(((0,2,0),))
pickedEdges =(ez0[0], ez1[0], ez2[0], ez3[0], ez4[0], ez5[0],
ez6[0], ez7[0], ez8[0], ez9[0], ez10[0], ez11[0])
p.PartitionCellByExtrudeEdge(line=ez12[0], cells=pickedCells, edges=pickedEdges,
sense=REVERSE)

```



**Fig. 1.** The side view of the model



**Fig. 2.** The whole pit foundation model

The Python script has used a number of cycle judgment statements to consider various soil conditions and different pit foundation excavation conditions. Through execution of the loops statement, the various horizontal soil layers are created and the retaining wall is created. Fig.1 shows the side view of the model which is drawing the horizontal soil layers lines in sketch. The whole pit foundation model is shown in Fig.2.

## 4 Material Assignment

After the soil and the retaining wall parts are made, the material property and section property of the soil and the retaining wall parts could be defined. An excerpt from the Python script is shown below:

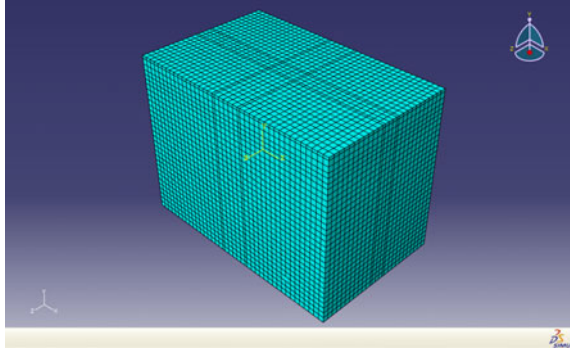
```

session.viewports['Viewport:1'].partDisplay.setValues(sectionAssignments=ON,engineeringFeatures=ON)
session.viewports['Viewport:1'].view.setValues(nearPlane=165.866,
farPlane=307.46, width=143.611, height=81.0078,
viewOffsetX=11.5043,viewOffsetY=0.715258)
p = mdb.models['Model-1'].parts['Part-1']
region = p.sets['outside soil']
p = mdb.models['Model-1'].parts['Part-1']
p.SectionAssignment(region=region, sectionName='Section-soil',
offset=0.0,offsetType=MIDDLE_SURFACE, offsetField=")
p = mdb.models['Model-1'].parts['Part-1']
region = p.sets['inner soil']
p = mdb.models['Model-1'].parts['Part-1']
p.SectionAssignment(region=region, sectionName='Section-soil', offset=0.0,
offsetType=MIDDLE_SURFACE, offsetField=")
p = mdb.models['Model-1'].parts['Part-1']
region = p.sets['wall']
p = mdb.models['Model-1'].parts['Part-1']
p.SectionAssignment(region=region, sectionName='Section-wall', offset=0.0,
offsetType=MIDDLE_SURFACE, offsetField=")

```

## 5 Meshing the Model

Now the structured meshing technique and adaptive meshing technique can be used to mesh the whole foundation pit model. The meshing results is shown in figure 3. An excerpt for meshing the model is shown below:



**Fig. 3.** The meshed foundation pit model

```
p = mdb.models['Model-1'].parts['Part-1']
session.viewports['Viewport: 1'].setValues(displayedObject=p)
session.viewports['Viewport: 1'].partDisplay.setValues(sectionAssignments=OFF,
    engineeringFeatures=OFF, mesh=ON)
session.viewports['Viewport: 1'].partDisplay.meshOptions.setValues( meshTechnique=ON)
p = mdb.models['Model-1'].parts['Part-1']
e = p.edges
pickedEdges = e.getSequenceFromMask(mask=('[#ffffff:19 #ffffff ]', ), )
p.seedEdgeBySize(edges=pickedEdges, size=2.0)
p = mdb.models['Model-1'].parts['Part-1']
p.generateMesh()
```

## 6 Establishment of the Boundary Conditions and Analysis Steps

The side faces are constrained in horizontal direction and the bottom face is constrained in vertical direction. And the analysis steps are created according to the foundation pit excavation stages.

Now the work of modeling foundation pit has been completed. This Python script could not only be used for modelling this JieFang Bridge foundation pit, but also be used for modelling other foundation pit. So the work of analyzing foundation pit would become easy by using the Python script presented by this paper.

## 7 Conclusions

(1) In ABAQUS, establishment and analysis of models are mainly divided into several parts: parts, assembly, property, mesh, interaction, step, submit and so on. For some complicated models, direct interactive modeling is very troublesome, and the efficiency is very low. So a Python script is presented to produce an input file which is compatible with the ABAQUS in order to make the foundation pit model easier and faster.

(2) According to the actual excavation stages of foundation pit, this paper has established a Python script which is suit to any soil layer and any excavation working procedure. The soil and the retaining wall parts could be automatically made, the material property and section property of the soil and the retaining wall parts could be automatically defined, the model could be automatically meshed, and the boundary conditions and analysis steps could be automatically defined. So the work of analyzing foundation pit would become easy by using the Python script presented by this paper.

**Acknowledgments.** This paper is the partial result of a research project sponsored by the soft science research project of the Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). The project NO. 2011-K3-32.

## References

1. Lu, Y.-N., Shen, L., Yue, J.-Y.: Analysis and performance for deformation and internal-force of retaining and bracing system of super-large deep excavation combined with substructure of main buildin. *Chinese Journal of Geotechnical Engineering*, 1365–1369 (November 2006)
2. Guo, H.-Z., Zhang, Q.-H., Zhu, J.-W., Ji, H.-M.: Application of soil coupled creep model to simulate foundation pit excavation. *Rock and Soil Mechanics*, 688–692 (March 2009)
3. Sun, J.-S., Dou, Y.-M., Liu, X.-Y.: Influence Analysis Of Soil Water Content on the Retaining Structure Bearing Characteristics of Foundation Pit, 91–94 (October 2008)
4. Zhang, R.-L., Xu, N.-W.: Numerical Simulation of Timbering Design of Deep Foundation Pits Based on PLAXI, 131–136 (2010)
5. Jin, Y.-Y., Bai, Y.: Elastic-plastic analysis for pile-anchor supporting system of deep foundation pit, 236–240 (2009)
6. Ying, H.-W., Guo, Y.: 3D analysis on a deep beam-slab braced foundation pit considering effect of construction process, 1670–1675 (2007)
7. ABAQUS inc. ABAQUS Analysis User's Manual, pp. 5–10 (2004)
8. ABAQUS inc. ABAQUS Scripting User's Manual, pp. 12–22 (2004)

# The Research of the Cloud Security Architecture

Hu Xiangyi, Ma Zhanguo, and Liu Yu

Beijing Municipal Institute of Science & Technology Information  
Beijing Key Laboratory of Network Cryptography Authentication  
Beijing, China

huxy368@sohu.com, mzgsy@163.com, lvden12121212@gmail.com

**Abstract.** A new solution of cloud security is proposed in this paper. It is used the security architecture of cloud computing based on symmetric encryption algorithm, combination key technology and encryption chip hardware. Compared with other cloud security solution, there are some advantages. The security level is higher. The security function is stronger. It is ensured the data is security and integrity transmission. It is guaranteed the security of program running and data storage of client. The real name and rapid login of user is realized, at the same time, the construction and maintenance costs of cloud security system are reduced.

**Keywords:** integrity verification, login log, combination key algorithm, symmetric encryption algorithm, gatekeeper.

## 1 Introduction

Currently, the cloud computing national development plans are formulated in all advanced countries. The cloud computing is the main battlefield that the world powers struggle the new strategic industries. The security of cloud computing is caused extensive attention by related researcher and security department. It is risen the study of cloud security technology. However, the cloud security system is created with the asymmetric cryptographic algorithm and symmetric cryptographic algorithm in the most cloud security solution. The cloud security certification system based on asymmetric cryptographic algorithm manages user amount less, the security certification and the login speed of the cloud user is slower, especially, the integrity verification efficiency of the data transmission is very low but the construction and maintenance costs are higher. Therefore the solution of only using symmetric cryptographic algorithm, combination key technology and encryption chip hardware are used to establish the cloud security system is proposed in this paper. The security level of the cloud security system is improved. The efficiency of data transmission integrity between the cloud client and the cloud computing platform is also improved. The large number of cloud user security and rapid login are implemented. The key management question only using the symmetric cryptographic algorithm is resolved. The front and back architectures of cloud computing center are built and linked by gatekeeper. The files of the cloud users are interacted by the file ferry function of the gatekeeper. Therefore, the security of the program running and the data storage of

client are guaranteed, at the same time, the construction and maintenance costs of cloud security system are greatly reduced.

## 2 Cloud Security Architecture and It's Security Solution

Cloud security architecture and security solution are building the cloud security center on the cloud computing platform. The center contains the authentication system, data integrity verification system, key management system, encryption/ decryption system and log management system.

1. The security architecture of the cloud computing is that the cloud security center is built on the cloud computing platform. The cloud security center connected respectively with the WEB server of cloud computing, the front-end server and the back-end server of the computing center. The cloud security center is responsible for the authentication of the cloud user or cloud computing manager to login WEB, the decryption and data integrity verification for the front-end and the back-end ciphertext file of computing center, the digital signature and file encryption for the back-end plaintext file of computing center. It is shown in figure 1.

2. The cloud computing platform contains two parts. One is the front-end of the cloud computing center. The other is the back-end. The front-end is connected with the cloud computing WEB server. It is connected between the front-end and the back-end with one gatekeeper equipment. The ciphertext file is ferried to the computing center back-end by the data ferryboat function of the gatekeeper and then the data ciphertext file results that the program of the computing center back-end computed are ferried to the computing center front-end when the program running in the back-end is required by the cloud user.

3. The intelligence chip E-KEY is used to create the encryption system at the client of the cloud user. A symmetric cryptographic algorithm, a symmetric key generation algorithm and a group of key seed data are written in the E-KEY. The key seed data of each cloud user are different from each other. The authentication protocol, the digital signature protocol, the signature verification protocol and the data encryption/ decryption protocol of the client are built in the E-KEY.

4. A symmetric cryptographic algorithm, a symmetric key generation algorithm and a set of fixed symmetric key K (memory key) are written in the encryption card chip of the cloud security center, at the same time, the authentication protocol, the digital signature protocol, the signature verification protocol and the data encryption/ decryption protocol of the cloud computing end are built in the chip of the encryption card. The identities of the all cloud users are saved in the database of the cloud security center. The identities are corresponded with a group of key seed data ciphertext. All the key seed data of the cloud users are respectively encrypted into ciphertext with the fixed symmetric encryption key K (memory key) during system initialization.

5. The cloud security center is composed by the server (or IPC) and the encryption chip hardware. The encryption chip is looked as the "arbiter" of the cloud user authentication and signature verification. The authentication and signature protocols are established merely using the symmetric cryptographic algorithm. Thus, the update and maintenance problem of symmetric key is solved by using the

combination symmetric key generation algorithm. Then the fast authentication protocol and the signature verification protocol are established.

6. On the client of the cloud user the file is encrypted into ciphertext first by the cloud user's E-KEY device encryption system in the client and then uploaded to the front-end of the cloud computing center. The ciphertext file computed by the cloud computing platform and received by the client is decrypted and verified the data integrity in the E-KEY device. It is shown in the file encryption/ decryption transmission topology figure 2.

7. The log database is built in the cloud security to record that the cloud user and the cloud computing administrators login the cloud computing platform and their operation.

7.1 The login log management function is created in the cloud security system. The login parameters of the cloud users and the cloud computing administrators login the cloud computing platform are stored in the login log database. The login parameters contain the identity of the cloud users or the cloud computing administrators, the timestamps and the random numbers (used to generate the data of authentication password), etc.

7.2 The operating log management function is established in the cloud security system. The parameters of the cloud users and the cloud computing administrators "clicking on" the cloud user's files (including: browse, modify, copy, send or print files and so on) are recorded in the operation log database. The parameters include: the identities of the cloud users and the cloud computing administrators, the names of the files, the timestamp, the random number, the digital signature and the client's IP address, etc.

### 3 The Security Protocols of the Cloud Security Center

The protocols of the authentication module, the digital signature protocols, the data integrity verification protocols, the data confidentiality protocols and the log management system are established in the cloud security center.

#### 1. The authentication protocol

The process of the cloud user login the WEB server of the cloud computing platform is two steps. Firstly, the user is authenticated by the client hardware device (E-KEY). The security center is responsible for certificating the identity of the user. That is, a set of random numbers and a timestamp are generated in the WEB server and then sent into the client E-KEY. The authentication key and authentication password 1 is generated based on the symmetric key generation algorithm composing by the timestamp, the random numbers in the E-KEY. Secondly, the user's identity, the timestamp, the random numbers and the authentication password are sent to the cloud security center. The authentication password 2 is generated in the encryption chip of the cloud security center. The identity of the cloud user is determined by comparing the authentication password 1 and 2. The legitimate user login the cloud computing platform and enter the work area corresponding with the cloud user. The illegal user isn't allowed login the cloud computing platform. The cloud computing platform includes front and back end of the computing center. The work area of each cloud user also includes two parts, front and back work areas.



2. The data integrity authentication and the data confidentiality transmission protocol.

2.1 After the legitimate user logs in the cloud computing platform and enters his work area a set of timestamps and random numbers are generated in the client side. The signature key is generated according to the symmetric key generation algorithm. The files (program and data) to be sent to the cloud computing platform are signed digitally and encrypted into ciphertext. Then the ciphertext (including the ciphertext of the file and digital signature), the identity of the cloud user, the timestamp, and the random numbers are sent to the front work area corresponding to the cloud user. The ciphertext files of all cloud users are stored in the computing center front-end of the cloud computing platform.

2.2 The program is directly run in the front-end of the computing center when the man-machine dialogue is demanded during the cloud users running the program. After the program running button is clicked by the user the process is as follows. Firstly, the ciphertext file will be run is decrypted by the encryption/ decryption system in the cloud security center and is stored in the plaintext file temporary storage area of corresponding with user workspace. Secondly, the data integrity is verified. If the file isn't pass the integrity verification the result of "the file data is incorrect, re-submit!" is returned to the user work area of corresponding the computing center's front-end. Otherwise, the program is run and the results (data) are obtained by the front-end of the computing center. The results are stored in the plaintext file temporary storage area of corresponding user workspace. The data is signed digitally and encrypted by the encryption/ decryption system in the cloud security center. If the results, that is, the plaintext data are demanded by the cloud user. Then the digital signature and the ciphertext are transmitted to the client of the cloud user. The received ciphertext is decrypted and data integrity verified using the E-KEY device in the client and then is generated the plaintext. The plaintext is readable. In addition, the plaintext files stored in the cloud user work area of the cloud computing center is determined by the user whether the files are reserved. The files are cleared if the user doesn't want the file is reserved and clicks the delete button. The security level of the protocol isn't high enough.

2.3 The program is run in the back-end of computing center when the man-machine dialogue is not demanded during the cloud users running the program. After the submit button is clicked directly by the user the process is as follows. Firstly, the corresponding ciphertext file (program and data) is ferried to the back-end of the cloud security center by the gatekeeper. Secondly, ciphertext received is decrypted and data integrity verified by the encryption/ decryption system. If the file isn't pass the integrity verification the result of "the file data is incorrect, re-submit!" is returned to the user work area of corresponding with the computing center's front-end. Otherwise, the program is run and the results (data) are obtained. The results are signed digitally and encrypted by the encryption/ decryption system. The digital signature and the ciphertext are ferried to the front-end of the computing center. That is, the file sent by the back-end of the computing center and then received by the front-end is ciphertext. The security level of the protocol is higher.

2.4 The authentication is passed and the cloud user logs in the work area in the front-end of the computing center. The ciphertext, that is, the results of the program running are downloaded to the client of the cloud user. The received ciphertext is

decrypted and data integrity verified using the E-KEY device in the client. It is shown in the figure 2.

### 3. The key management protocol of the symmetric algorithm

3.1 A set of random numbers and a timestamp are generated in the E-KEY device of the cloud user client. A set of key seed data are selected by the random numbers and the timestamp based on the symmetric key generation algorithm and generated a set of symmetric key. The key is treated as the encryption key (including: authentication, signature and encryption).

3.2 The key seed ciphertext is located in the key seed database by the cloud security center base on the corresponding cloud user's identity. At the same time, a set of random numbers and a timestamp are generated. The key stored in the encryption chip of the cloud security center is called and the set of key seeds are decrypted and generated plaintext files. The set of key seed (plaintext file) is selected based on the symmetric key generation algorithm composed by the set of random numbers and the timestamp and generated a set of symmetric key. The key is treated as the encryption key (including: authentication, signature and data encryption).

### 4. The log management protocol

4.1 The cloud user or the administrator using the E-KEY device pass the authentication and login the cloud computing center and then login the log database management system. The login parameters of the cloud user or the administrator login the cloud computing platform are stored in the login log database. Where, the login parameters contain the identity of the cloud user or administrator, the timestamp and the random numbers (used to generate the password data), etc.

4.2 The file name (including: browse, modify, copy, send or print files and so on) is clicked after the cloud user or the administrator login the cloud computing center. The parameters of operating the cloud user's files by the user or the administrator are stored in the operation log management system. Where, the login parameters contain the identity of the cloud user or administrator, the files name, the timestamp and the random numbers, digital signature and the client's IP address, etc.

## 4 The Advantages of the Cloud Security Solutions

The cloud security architecture established by the alone symmetric encryption algorithm, the combination key algorithm and encryption chip device have a greater advantage.

1. Scale user login is meted based on the authentication protocol. It is faster. Through the test by third party authorities department the speed is faster and the efficiency is higher of the authentication and the signature verification. It is concurrent certification 20,000 person/ 27 second. It is concurrent validation 20,000 / 46 second.

2. The efficiency of the security protocols is high and the construction costs are low. The task of 18 million security protocols (authentication, decryption and signature verification) are completed through the cloud security center built by this technology on the information security system practical application of the Internet of things Beijing demonstration project. And 3 million sensor equipments are managed

on the application. The construction costs of the cloud security center are significant saved due to the obvious high efficiency of the security protocols.

3. The authentication protocol, the digital signature protocol and the data encryption/ decryption protocols are implemented in the chip. They are the “chip level” security protocols. Comparing with the existing authentication and digital signature protocol based on asymmetric cryptographic algorithm safety performance of this paper’s solution is much higher. The symmetric key (for authentication, signature and data encryption) is real-time generated using the combination symmetric key generation algorithm. The symmetric is changed each time and not repeated. The safety level of the cloud security system is effectively improved.

4. The files submitted by the all cloud users are stored in the front-end of the cloud computing center according to the work area assigning of rights management system. The files are encrypted to the ciphertext before the cloud users submit. Thus, it is prevent the hackers to get users’ files (including: program and data) through the virus program. It is effectively ensure that the user’s file storage safety.

5. The file submitted to he front-end of the cloud computing center by the cloud user is ciphertext. The file transmitted between front-end and back-end of the cloud computing center is ciphertext. And the ciphertexts are exchanged through the gatekeeper’s ferry function. Thus, it is ensure that the files (including: program, data and results) transmitted from the cloud user client to the back-end of the cloud computing center through the front-end are transfer security. At the same time, the virus invasion is blocked by the gatekeeper device.

6. The whole operation process of the cloud users and the cloud computing administrator is tracked by the log management module. It is prevent the files of the legitimate users are acquired and damaged by the illegal users (the outside ghost) and the internal manager (inside God). It is shown in figure 3.

## 5 Conclusions

The cloud security solution is that the cloud security system is built around the cloud computing platform. It is using the proven technology, that is, the symmetric cryptographic algorithm, combination key technology and the smart chip technology. The cloud security architecture contains the user authentication, the rights management, the data confidential transmission, the data integrity verification, the secure storage and the log management, especially using the gatekeeper connecting the front-end and the back-end of the cloud computing center. The ciphertext file is decrypted in the back-end of the cloud computing center. The program is also run in the back-end. The virus attack chain is effectively blocked by the gatekeeper. The probability of the plaintext acquired or damaged by the virus is greatly reduce. The user authentication, the data integrity verification, the key management and other core protocols involved by this solution are won the security review of our country’s relevant industry. The protocols are safety and reliable. And the protocols are tested by the national software assessment center and passed. Each technical indicator is ahead of the traditional cloud security system using the asymmetric cryptographic algorithms.

## References

1. Hu, X.: A safe and effective way of network user authentication, China, 200610103357.8 (September 9, 2009)
2. Hu, X.: A method of building the VPN system based on dynamic encryption algorithm, China, 200610144395.8 (August 1, 2008)
3. Hu, X.: The digital signature method based on the combination key, China, 200810113101.4 (August 1, 2011)
4. Hu, X., Zhao, G.: A CSK-based Solution for Person Authentication. In: The Seventh Wuhan International Conference on E-Business: Unlocking the Full Potential of Global Technology, pp. 244–249 (2008)
5. Hu, X., Li, Y.: The Use of Symmetric Cipher in Identification. *Network Security Technology & Application* (3) (2007) (in Chinese)
6. Nan, X.: 2005 Information Security Step into the “trust” times. *The World of Computer* (2005) (in Chinese)
7. Hu, X., Zhao, G., Xu, G.: Security Scheme for Online Banking Based on Secret Key Encryption. In: The Second International Workshop on Knowledge Discovery and Data Mining, Moscow, Russia, pp. 632–635 (2009)
8. Zhao, G., Hu, X., Li, Y., Du, L.: Implementation and Testing of an Identity-based Authentication System. In: ISECS International Colloquium on Computing, Communication, Control, and Management, pp. 424–427 (2009)
9. Du, L., Hu, X., Li, Y., Zhao, G.: A CSK Based SSL HANDSHAKE PROTOCOL. In: The IEEE International Conference on Network Infrastructure and Digital Content (2009)

# The Energy Harvesting Tipping Point for Wireless Sensor Applications

Alan Pang<sup>1</sup> and Ross Bannatyne<sup>2</sup>

<sup>1</sup> Asia-Pacific MCU Senior Marketing Manager, Silicon Laboratories

<sup>2</sup> MCU Marketing Manager, Silicon Laboratories

**Abstract.** In recent years, energy harvesting technology has become more sophisticated and efficient, and energy storage technologies, such as supercapacitors and thin-film batteries (TFBs), have become more cost-effective. In addition, sophisticated low-power integrated circuits are now available to perform useful functions in energy harvesting applications, such as algorithmic control and wireless communications using tiny amounts of energy. We have now reached a technological tipping point that will result in the evolution of energy-harvesting-based systems from today's niche products, such as calculators and wrist watches, to their widespread deployment in building automation, security systems, embedded controls, agriculture, infrastructure monitoring, asset management and medical monitoring systems.

The wireless sensor node is one of the most important product types poised for growth as an energy-harvesting solution. Wireless sensors are ubiquitous and attractive products to implement using harvested energy. Running mains power to wireless sensors is often neither possible nor convenient, and, since wireless sensor nodes are commonly placed in hard-to-reach locations, changing batteries regularly can be costly and inconvenient. This presentation will discuss how to implement wireless sensors powered by harvested energy coupled with ultra-low-power, single-chip wireless microcontrollers (MCUs) that use sophisticated power management techniques.

**Keywords:** Energy Harvesting, Wireless Sensor Node, Wireless MCU, Solar Energy, Ultra-low Power.

## 1 Introduction

As wireless sensor nodes become commonplace in building automation, security systems, embedded controls, agricultural technology, infrastructure monitoring, asset management and medical monitoring systems, it makes sense to power these ubiquitous sensor networks with economical and reliable sources of energy. In many cases, a primary battery cell provides a simple, low-cost power source for a wireless sensor node.

However, there are many examples in which an energy harvesting system provides a more cost-effective alternative for wireless sensor nodes than batteries or mains power supplies. Running mains power to wireless sensors is often neither possible nor convenient, and, since wireless sensor nodes are commonly placed in hard-to-reach locations, changing batteries regularly can be costly and inconvenient.

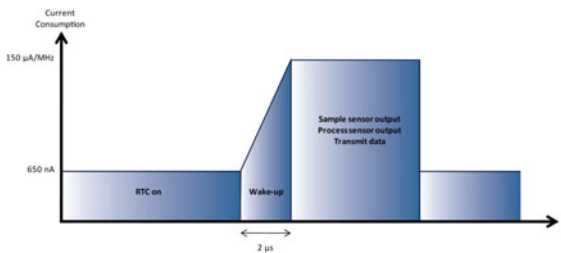
In some applications, battery replacement cost can be orders of magnitude greater than the cost of the battery itself or even the cost of a self-sustaining energy harvesting system. For example, a simple service call to replace a battery in a wireless utility meter may cost hundreds of dollars. Or consider wireless sensor-based infrastructure monitoring systems. Sensors buried inside suspension bridges and highway overpasses provide access to a tremendous amount of mechanical energy (through the constant vibration of passing vehicles), which can make energy harvesting much more attractive than traditional batteries in these applications, given the challenging physical environments and the cost of servicing the sensor nodes.

In recent years, advances in energy storage technologies such as supercapacitors and thin-film batteries (TFBs) have provided cost-effective building blocks for energy harvesting systems. In addition, mixed-signal devices that can perform useful functions, such as algorithmic control and wireless communications, have become more cost-effective and exceptionally power-efficient. We have now reached a technological tipping point that is driving the evolution of energy-harvesting-based systems from today's niche products, such as calculators and wrist watches, to their widespread deployment in wireless sensor nodes.

Although systems powered by harvested energy sources have existed for many years, developers have found it challenging to implement wireless sensor nodes within very low power budgets. The off-the-shelf availability of ultra-low-power wireless microcontrollers (MCUs) and transceivers, coupled with thin-film battery and solar cell technologies, has now made it feasible to design self-sustaining wireless sensor nodes powered by harvested energy sources.

## 2 Optimizing Wireless Systems for Ultra-Low Power

Low-power modes on MCUs and wireless transceivers have been optimized in recent years to enable effective power management in wireless sensor applications. Figure 1 illustrates a typical wireless sensor node power cycle.



**Fig. 1.** Wireless Sensor Node Power Cycle

The designer's goal is to minimize the area under the curve in Figure 1, which corresponds to power consumption. Power consumption can be minimized by optimizing the relative amount of time spent in low-power sleep mode and reducing the active mode time. A fast processing core enables the MCU to execute the control algorithm very quickly, enabling a rapid return to low-power sleep mode and thereby minimizing the power-hungry area under the curve.

Wireless sensor nodes spend most of their time in sleep mode. The only subsystem that stays awake is the real-time clock (RTC). The RTC keeps time and wakes up the wireless sensor node to measure a sensor input. Low-power RTCs typically integrated onto microcontrollers consume only a few hundred nanoamps. It is important to minimize the system's wake-up time because power is consumed during this time. An RTC uses a free-running counter in the MCU timer subsystem. When the free-running counter rolls over, it generates an interrupt that wakes up the MCU often. If a 32.768 kHz crystal is used, a 16-bit free-running counter rolls over every two seconds and wakes up the MCU. If a wider free-running counter, such as a 32-bit counter, is used, the periodic interrupt occurs less often, and additional power may be conserved.

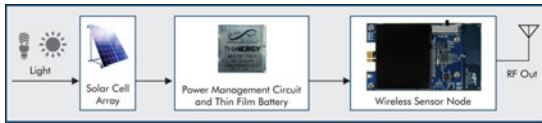
When a wireless sensor node wakes up, it is usually intended to measure a sensor signal using the analog-to-digital converter (ADC). It is important to note the wake-up time of the ADC as well as the digital wake-up time since there is little point in waking up the CPU very quickly if the ADC takes an order of magnitude longer to wake up. A low-power MCU should wake up both the CPU and the ADC in a couple of microseconds. When the sensor node is awake, the MCU current is typically approximately 160  $\mu\text{A}/\text{MHz}$ . When the sensor data has been measured, the algorithm running in the MCU decides whether the data should be transmitted by the radio. To send the data, a low-power ISM band radio consumes somewhat less than 30 mA for only a millisecond or so. When this peak current is averaged out, the overall average current consumption of the wireless sensor node is in the low microampere range.

The radio transmission consumes most of the current in the system. Minimizing the amount of time the radio is on is essential to conserving energy. One way to achieve this is to avoid complicated communications protocols that require the transmission of many bits of data. Steering clear of standards with large protocol overhead is desirable when power is at a premium. It is also important to consider the desired range. Wireless range can be traded for power consumption. An interesting approach to balancing this trade-off is to use dynamic ranging, which allows full-power transmissions when maximum energy is available but reduces the output power level when harvested energy is limited.

Another way to reduce the wireless sensor node's power consumption is to minimize the number of chips used in the system. Fewer chips on the printed circuit board (PCB) result in lower leakage current losses. Using an MCU that integrates as many functions as possible ultimately helps reduce overall current consumption. If a dc-dc converter is integrated onto the MCU, it can be switched off when the MCU is sleeping. Silicon Labs' Si10xx wireless MCU, for example, contains an integrated dc-dc converter that allows the system to be powered by a single AAA alkaline battery and still achieve 13 dB output power at the antenna. Because of its high level of integration and ultra-low power consumption, the Si10xx wireless MCUs have been used successfully in energy harvesting wireless sensor nodes.

### 3 Managing Harvested Energy

An important consideration in the development of an energy harvesting sensor node is to ensure that there is always enough energy available to power the system, as shown in Figure 2.



**Fig. 2.** Typical Energy Harvesting System

This energy harvesting system uses a solar cell array to harvest energy. A solar cell unit, such as a Sanyo AM-1815, delivers approximately  $40\ \mu\text{A}$  when a  $200\ \text{Lx}$  light level is available. It is reasonable to expect this level of light in an office with a window but no direct sunlight on the cell. The  $40\ \mu\text{A}$  of current that the array generates is fed into a power management circuit and trickle-charged into a thin-film battery (TFB). When selecting power management chips, it is necessary to pay attention to the leakage current characteristics, which are normally only a few microamperes. However, with only  $40\ \mu\text{A}$  coming into the TFB, this tiny amount of leakage must be understood and accounted for. A thin-film battery, such as the Infinite Power Solutions MEC101-7SES, provides a  $0.7\ \text{mAh}$  capacity, which is a reasonable amount of energy for a wireless sensor node system. At a  $200\ \text{Lx}$  level of light, this TFB charges up fully in around 17.5 hours.

This combination of solar cell, power management and storage technologies provides an adequate level of energy for a wireless sensor node. The next important decision in the design process is the selection of a low-power MCU and wireless transceiver combination that can operate effectively from a limited energy source. A Silicon Labs Si1012 wireless MCU is an ideal choice because of its extremely low power consumption and high-performance radio characteristics. This wireless MCU uses a programmable sub-GHz ISM radio in a single-chip configuration with an ultra-low-power MCU; the highly-integrated device, which also includes an on-chip temperature sensor, is essentially a wireless sensor node on a chip.

With the hardware configured as shown in Figure 2, the control problem to be considered is how to operate the wireless sensor node at a duty ratio that does not deplete the TFB capacity that is itself being trickle-charged by the solar cell. Using the low-power design techniques discussed earlier, it is possible to reduce the average current of the wireless sensor node to around  $51\ \mu\text{A}$  (including power management leakage) while transmitting sensor data every second for three minutes. That is low enough to allow the system to operate and stay fully charged in minimal lighting conditions.

If the light input is reduced to  $0\ \text{Lx}$ , the wireless sensor node continues to operate and transmit for 64 hours before the TFB capacity is exhausted (assuming the three-minute transmit period is repeated every 20 minutes). A simple spreadsheet detailing expected input energy (i.e. how much light is available) versus output energy (how often the node is required to transmit) is the only tool that a designer needs to optimize the system. If more than adequate light is expected, this energy can be used to increase the range of the transmitter. This type of system allows a range up to 300 feet, depending on the exact conditions.

Many different types of energy harvesting sources can be used to power a wireless sensor node instead of using a solar cell (or even in combination with solar energy). If a wireless sensor node is placed in a location without ready access to a light source,



the node can be powered by thermal, vibration (piezoelectric) or radio wave energy harvesting sources. The power management, storage and wireless sensor node circuits are essentially the same as those used in the solar cell example. Regardless of the harvested energy source, the system design principles are the same: a limited source of energy is captured and stored in a TFB and then used to power an ultra-low-power wireless sensor node.

## **4 Enabling “Green” Wireless Sensor Nodes**

The ability to power wireless sensor nodes from harvested energy sources allows embedded designers to offer wireless networking systems with significantly reduced cost of ownership for the end-user as well as green benefits for the environment, with fewer batteries ending up in landfills. The cost of replacing batteries housed in out-of-the-way sensor node locations can be quite significant. These wireless sensor nodes, for example, can be embedded in structures, such as buildings or bridges, or even buried underground.

The three key enabling technologies needed to create self-sustaining wireless sensor nodes are readily available today: cost-effective energy harvester devices, small and efficient energy storage devices and single-chip ultra-low-power wireless MCUs. Wireless sensor nodes powered by harvested energy sources will soon become commercially viable and commonplace technologies used in our homes, offices, factories and infrastructure.

# Design of Broadband Optical Access Network Based on 10G-EPON

Duo Peng<sup>1</sup> and Jianbiao Wang<sup>2</sup>

<sup>1</sup> Computer and Communication College, Lanzhou University of Technology  
Lanzhou, China 730050

<sup>2</sup> Engineering Department, Lanzhou Lianzhi Software Development Co., Ltd.  
Lanzhou, China 730000

{pengduo7642, lzwangjb}@163.com

**Abstract.** 10G-EPON is the latest generation of access network technology, this paper describes the system structure and technical characteristics, analysis applications of 10G-EPON. A typical application environment as the background, given the overall solution 10G-EPON program, design and analysis of the FTTB + DSL access network architecture and optical power budget. It implemented smooth upgrade from 1G-EPON to 10G-EPON and with high feasibility and superiority.

**Keywords:** 10G-EPON, optical power budget, access network.

## 1 Introduction

At present, the passive optical network (PON) technology has become the main access means of broadband optical access. It has many features: save the trunk optical fiber, use fewer optical interfaces of the central office equipment, greater service area, flexible bandwidth allocation, costs low of operation and maintenance. In it, EPON which has become the preferred broadband optical access technology. It has a lot of applications in the broadband access network construction and Fiber optic cable instead of copper access network transfor-mation process [1].

Network has become an indispensable part of people's lives, users' access the bandwidth requirements increase dramatically. Although EPON able to provide 1Gbps downstream bandwidth, but with some new services (such as HDTV, online games, etc.) became more common, which put forward higher requirements to PON technology. Such as video broadcasting and VoD from the DTV transition to HDTV, the bandwidth required for each program 20 Mbps (MPEG-2) or 10 Mbps (H.264). Per-user bandwidth demand expected to grow at an order of magnitude every five years and has trend of accelerated, so the IEEE, in September 2009, enacted a new generation PON access technology standard IEEE802.3av TM-2009, the 10G-EPON standard. As the technical characteristics of the EPON network of point-multipoint and multi-service bearing capability makes the 10G-EPON networking and planning work is facing many new challenges.

## 2 10g-Epon Architecture and Technical Characteristics

EPON is a broadband access technology which is offers a variety of integrated services, using passive optical transmission, high-speed Ethernet-based platform and the TDM (Time Division Multiplexing) media access control method [2]. 802.3avTM-2009, also known as Physical Layer Specifications and Management Parameters for 10 Gb /s Passive Optical Networks. It make the downstream/upstream bandwidth from 1G to 10time (up to 10Gbps), and Compatible with 1G-EPON solutions of network protocols and topologies. 10G-EPON standard provides two rate models, 10Gbps downstream, 1Gbps upstream of the asymmetric mode (10/1GBASE-PRX) and 10Gbps uplink and downlink symmetric mode (10GBASE-PR). Asymmetric mode can be considered a transitional form symmetrical pattern, in the early demand for upstream bandwidth less and cost-sensitive applications can use the non-symmetrical mode. As the business development and technological progress, it will gradually transition to the symmetric mode. 10G-EPON wavelength assignment is shown in Figure 1. Downstream using 1574-1580nm band, upstream using the 1260-1280nm band and overlap with 1G-EPON wavelengths.

Thus, in the downstream direction, 10Gbps and 1Gbps signal using WDM mode. In the upstream direction can not be used WDM mode, only the use of dual-rate TDMA approach [3].

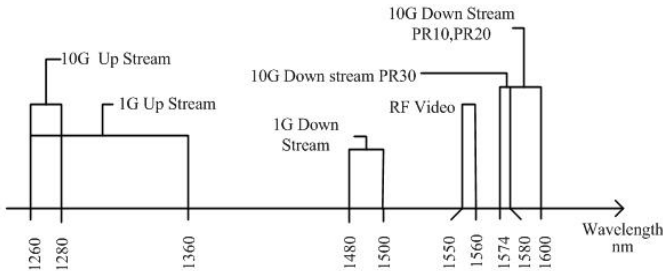


Fig. 1. System Wavelength of 10G-EPON and 1G-EPON

10G-EPON currently provides three kinds of power budget, which are PR10/PRX10, PR20/PRX20 and PR30/PRX30. Power budget of PR-symmetry type be shown in Table 1. It can be seen from the table of power budget 10G-EPON can support up to 20Km transmission distance and 1:32 split ratio [4, 5].

Table 1. 10G-EPON optical power budget

Basic Features	Distance (Km)	Splitting ratio	Insertion Loss (dB)	Power budge (dB)
PR10	10	1:16	~20	~22
PR20	20	1:16	~24	~26
	10	1:32	~24	~26
PR30	20	1:32	~29	~31

In 10G-EPON, the upstream and downstream data stream using different methods. When the OLT initialize, it periodically broadcasts the slot allow to access to its port and other information. After optical network unit (ONU) on power, according to information of allow access to by OLT broadcasts, initiate the registration request, certified by the OLT, ONU allowed access, and assign it a unique logical link identifier (LLID). From the OLT to multiple ONU (downstream direction) using broadcast to send data frames, each frame header includes the ONU-specific logical link identifier (LLID) that of distributed in front. Through 1: N passive splitter data frame arrives ONU, then ONU according LLID to determine in the physical layer, receive to its own data frame, get rid of those data to other ONU frames. Data frame can be to all ONU (broadcast) or a special group ONU (multicast) [6]. Upstream direction (from ONU to OLT) using time division mode (TDMA) to share the system band, System for each uplink data of ONU is assigned a specific time slot. Thus, when the data converge to a common optical fiber, each ONU's upstream package does not interfere with each other, the downlink/uplink data flow shown in Figure 2 and Figure 3.

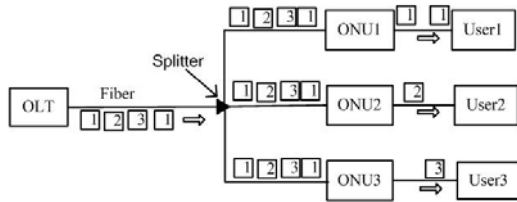


Fig. 2. EPON downstream data flow diagram

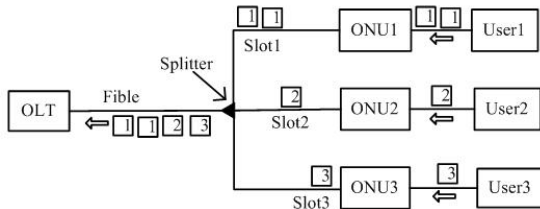


Fig. 3. EPON upstream data flow diagram

10G EPON using 64B/66B coding, efficiency of 97%, Compared with 1G-EPON (8B/10B efficiency of 80%) has improved significantly. FEC Functional of 10G-EPON use RS (255,223) coding, coding gain of 7.2dB, compared with 1G-EPON of RS (255,239) code stronger.

### 3 Applications with 10g-Epon

10G-EPON mainly used in two occasions. First, the community online via telephone line, no broadband access, at this time 10G-EPON used in PON-based access

technology to replace copper. In this new application, from beginning to deploy 10G-EPON network will reduce a lot of renovation costs for new residential and high-end users. Another application of the access area has been laid 1G-EPON but the bandwidth can not meet the user requirements, in this case 10G-EPON coexist with EPON and gradually replace the latter. Because from the 1G-EPON to 10G-EPON upgrade occurs gradually, must allow some users to go to the 10G-EPON, while the remaining users still stranded in EPON, the transfer from EPON to 10G-EPON may occur downlink in the first, then extended to the uplink. Because the wavelength of 10G-EPON is carefully designed, it will not interfere with existing 1G-EPON network, so allowing for a smooth upgrade from 1G-EPON to 10G-EPON [7, 8]. 10G-EPON and 1G-EPON mixed mode of operation shown in Figure 4.

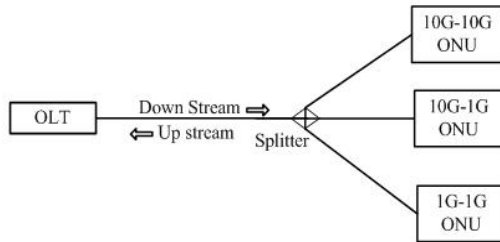


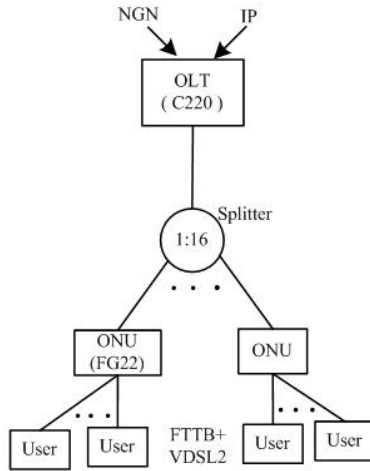
Fig. 4. 10G-EPON and 1G-EPON mixed mode

#### 4 Design of Broadband Optical Access Network Based on 10g-Epon

A typical design of broadband optical access network in urban residential area to illustrate the 10G-EPON-based access network design.

##### Selection and Analysis of Design

Because of the high-density residential district, users are more concentrated, and the network of high demand, should be way FTTx access. Through investigation that the area has been paved wired telephone lines, taking into account the full use of existing resources, from the cost, performance, consider using FTTB + DSL solutions. To meet the demand for HDTV and other high-speed services, ultimately determine the use of 10G-EPON + VDSL2 solution. VDSL2 has a relatively higher bit rate than ADSL2 +, it also downward compatible with ADSL2 + technology, users can meet the needs of high-definition and higher definition streaming services, the overall bandwidth of up to 18M or more. Broadband fiber access network structure is shown in Figure 5.



**Fig. 5.** Broadband fiber access network structure

For example, the district has 10 residential buildings, 500 households, distance the community communications room about 10Km. For ease of operation and management, reduce operating cost, the OLT placed in the central office switch room. Splitter can be placed ODF rack in the Community Center room, splitter ratio can be of 1:16, equal power spectral, and use optical cable via the branch connected to the corridor, ONU devices placed in residential buildings (usually in the basement). In user-intensive and network traffic area, increase the number of ONU, while reducing the connection number of users of ONU node. For users the relative dispersion (such as double floor, house) or the areas of less network traffic can reduce the number of ONU, while increasing the number of connected users of ONU node. ONU output signal through ODF and the twisted pair connection to the user's VDSL2 Modem.

### Optical Power Budget

ODN of the optical power budget ( $P$ ) is defined as the loss allowed by  $S / R$  and  $R / S$  ( $S$ : light emitting point of reference,  $R$ : optical receiver reference point) between the reference point of light loss, the use of units of dB. The loss includes the loss of fiber ( $a_f$ ) and passive optical fiber components (such as optical splitter loss ( $p_s$ ), active connector loss ( $a_c$ ) and splice loss ( $a_s$ ), etc.). ODN allowable loss values of downstream and upstream directions are the same. There are three main methods of Optical path loss calculations: the worst value method, statistical method and joint design method. In view of the access network environment, transmission distance is very short, usually without the use of joint design method, and suggested the worst value method, the calculation formula (1) as follows:

$$a_s + a_c N + a_f L + p_s < P \quad (1)$$

In this scenario, using the worst value method,  $a_f$  of 0.4dB/km,  $p_s$  to 12dB,  $a_c$  by 0.5dB/connector selection, there are four active connector ( $N=4$ ) from the OLT PON port to ONU PON port, total active connector loss of 2dB,  $a_s$  obtained 0.5dB,  $L$  for

the 20km, P is 26dB for the PR20 standard. Parameters into equation (1) can be obtained:

$$2\text{dB}+0.5\text{dB}+8\text{dB}+12=22.5\text{db}<26\text{dB} \quad (2)$$

By the formula (2) shows that, in this scenario the loss of ODN to meet the standards.

### Other Service Access

The 10G-EPON-based FTTB solutions in addition to for the broadband data services, but also voice, cable TV and other compatible service and expansion of integrated access.

In 10G-EPON + VDSL2 scenario, ONU connecting IPDSLAM or integrate DSLAM into the ONU directly , the use of existing telephone twisted pair transmission of voice signals and broadband data signals, users only need to install the DSL-Modem connect user's computer and Splitter connect ordinary telephone device, which is similar to using ADSL. For the traditional telephone service, it can provide POTS interfaces in the ONU, integrated V5 protocol interface in the OLT and access to the PSTN network. For analog and digital cable signals can be used to the method of increase a beam, the so-called "three-wave in single-fiber" WDM transmission. Separate CATV wavelength is 1550nm (shown in Figure 1). As the 10G-EPON systems and VDSL2 transmission rates are very faster, VDSL interface board can provide users with a variable transmission bandwidth up to 10 Mb/s. This program allows the user does not have to re-wiring, low cost, good performance.

10G-EPON systems have good expansibility, the program has been set aside for PON interfaces, when extend only add branch fiber from splitter to residential building ODF of new users. This way to complete the expansion of the network is easy and no impact on existing service.

### Upgrade from EPON to 10G-EPON

In the initial drafting, IEEE 802.3av standard was consider the 10G-EPON with 1G-EPON compatibility problems, in the wavelength distribution and multi-point control mechanism has a special consideration to ensure that 10G-EPON and 1G-EPON systems coexist on the same ODN. If the area has previously been deployed EPON access without changing the original FTTB topology and copper cables topology even power supply equipment, simply use 10G-EPON card replace OLT EPON line cards and ONU EPON uplink cards to complete 10 times the bandwidth enhance of smooth upgrade for end users. The user side for EPON + ADSL2 + users, just mixed plug VDSL2 board in the ONU and replace the ADSL2 + Modem with VDSL2 Modem . Upgrade block diagram shown in Figure 6. The whole upgrade process is the equipment board reform, without transformation of the most complex ODN network and equipment infrastructure, so upgrade is low cost.

In the control protocol and management mechanisms, 10G-EPON inherited the existing EPON standard, adapted part of the MAC layer, while expand the EPON MPCP protocol (IEEE802.3), increase of notification and consultation mechanism for 10Gbit/s capabilities. These measures can ensure 10G-EPON use of operation and management mechanism of EPON.

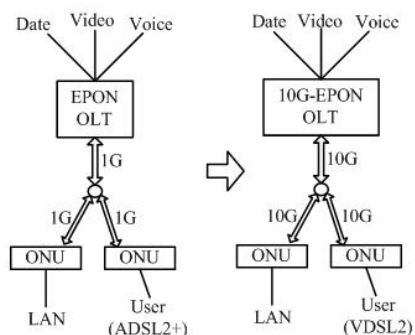


Fig. 6. From FTTB scenarios upgrade to 10G-EPON

## 5 Conclusion

This paper introduced the 10G-EPON technology features and system architecture, and described 10G-EPON scenarios. To Urban Residential District, for example, its access network has been designed, made using FTTB + DSL residential broadband service access solution to solve integrated service access problem of community. For the program selection, optical power budget, system upgrade to the detailed design. As the 10G-EPON multi-service access capability and VDSL high-bandwidth and multi-service support capabilities, the use of 10GEPON + VDSL2 technology is the ideal solution of high-speed broadband access for residential users.

As the 10G-EPON uplink uses TDMA mechanism in coordination to ensure that only one ONU working at the same time, which led to the time delay is relatively large for 10G-EPON uplink, especially negative for delay-sensitive services (such as VOIP). In 10G-EPON, uplink direction must to work in burst mode, therefore, it require higher indicators of the ONU burst mode optical transmitter and the OLT burst mode optical receiver. These problems hindered the popularity of 10G-EPON. With the continuous development of EPON technology, especially 10GEPON optical module technology matures, 10G-EPON technology will be an effective way to solve broadband access.

**Acknowledgments.** This work is supported by the Lanzhou University of Technology Natural Science general cross-cutting project, No: H1114cc001.

## References

1. Zheng, J., Mouftah, H.T.: Media access control for Ethernet passive optical networks: An overview. *IEEE Communications Magazine* 43, 145–150 (2005)
2. Kramer, G., Mukherjee, B., Pesavento, G.: IPACT: a dynamic protocol for an Ethernet PON (EPON). *IEEE Communications Magazine* 40(2), 74–80 (2002)
3. Chio, S.: Cyclic polling-based dynamic bandwidth allocation for differentiated classes of service in Ethernet passive optical networks. *Photonic Network Communications* 7(1), 87–96 (2004)



4. Kramer, G., Pesavento, G.: Ethernet passive optical network (EPON): Building a next generation optical access network. *IEEE Communications Magazine* 40, 66–73 (2002)
5. Assi, C., Yinghua, Y., Sudhir, D., et al.: Dynamic bandwidth allocation for quality of service over Ethernet PON. *IEEE Journal on Selected Areas in Communications* 21, 1467–1477 (2003)
6. Effenberger, F., Feng, D.N., Leung, R.: Consideration on the upstream burst structure. In: *IEEE 802.3 Interim Meeting*, Portland, OR, USA, January 22 (2008)
7. Banerjee, A., Park, Y., Clarke, F., et al.: Wavelength division multiplexed passive optical network (WDM-PON) technologies for broadband access: a review. *Journal of Optical Networking* 4, 737–758 (2005)
8. Naser, H., Mouftah, H.: A joint ONU interval based dynamics scheduling algorithm for Ethernet passive optical networks. *IEEE/ACM Transactions on Networking* 14(4), 889–899 (2006)

# Comprehending Collective Wisdom in Hall for Workshop of Metasynthetic Engineering Based on Semantic Web

Sao-qi Zhou, Ren Deng, and You-hong Fan

Logistical Engineering University

**Abstract.** In 1992, Prof Qian Xue-sen put forward the relevant theoretical framework of the handling of the "open complex giant system", namely, Hall for Workshop of Metasynthetic Engineering (HWME). The system solves complex problems by fully inspiring collective wisdom. Meanwhile, the whole system consists of the expert system, the machine system and the knowledge system. Therefore, how to realize the mutual communication and understanding between man and man, man and equipment, and man and system is the key to stimulate collective wisdom. This paper presents a semantic network as a bridge for communication between those different parties; analyses understanding process of collective process based on the semantic network, studies the expression method of semantic web knowledge on services and role speaking and finally clarifies formation of consensus of discussion based on the semantic network.

**Keywords:** Semantic Web, HWME, Collective Wisdom, Consensus of Discussion.

## 1 Introduction

At the beginning of 1990s', with the development of system science and computer technology, Prof Qian xue-sen put forward the theory of open complex giant system and its methodology, and thought that the effective solution is Hall for Workshop of Metasynthetic Engineering (HWME)[1-3]. HWME is a virtual workshop composed of experts system, machines system and knowledge system. It joins the role of experts, machines and knowledge, especially, emphasis on the roles of experts, which can enlighten or inspire each other to get the more advanced answers than any individual, that's collective wisdom.

The key problem in HWME is how to understand collective wisdom, and finally achieve the consensus of discussion [4-6]. Collective wisdom exists in process of mutual communication between experts and groups, and the understanding of collective wisdom goes through the whole procedure, involving all the parts in the discussion. In HWME, the communication between experts, experts and equipment, experts and networks needs understanding semantic from different part [7]. Therefore, with the knowledge limitation of individual part, each part should get an access to learn the new semantic words and vocabulary.

Semantic web is the bridge between the parts in HWME.

## 2 Introduce of Semantic Web

Semantic web is presented by Tim Berners Lee in XML2000 committee [8], and then the standards and technical solutions of semantic web are given by W3C org. RDF (Resource Description Framework), is one knowledge expression model among them, which presents a general resource description method. The basic construct of RDF is statement, which is a three elements set, including resource, attributes of resource and the attributes values.[9] One or many statements are constitute of description of ontology, so do the other a sets of ontologies, then construct the relationship networks of ontologies, then construct the whole virtual world. This semantic web, is a network graph expressed by semantic relationships between resources and attributes, which presented as [10-12]

$$SN = (O, P, f, \lambda) \tag{1}$$

$O$  is a resource or a set of resources,  $P$  is an attribute value or a set of attributes values, the elements in  $O, P$  are the nodes in semantic web;  $f$  is the set of attributes between  $O$  and  $P$ , it is mapping set from  $O$  to  $Q(P)$  ( $P$ 's power set),  $\{f \mid f : O \rightarrow Q(P)\}$ ;  $\lambda$  is partial order of  $O$ , in fact, it is *ISA* or *Part-of* relation arc in semantic web. For  $f_i \in f, i \leq |f|$ , means  $O$  has  $i$  attributes, and for  $f_i$ , it is pointed out by attribute direction arc in semantic web or calculated by attribute inherited.

Then, given the BNF description of semantic web [13]:

```

<Semantic web>:=<Resource>Merge(<Resource>,...)
<Resource>:=< Resource ><Semantic relations>(< Resource >, ...)
<Resource >:=(<Attribute-Value>, ...)
<Attribute—value >:=<Attribute >: <Value>
<Semantic relations>:=<Predefined Semantic relations >
<User defined Semantic relations>
    
```

In which, Merge(...) is a combination process, which combines all the resource and relationships in the brackets, and merge the same defined node to a unique one, then made the description and meaning of resource and its attributes is unique in current domain, so all the requirements of this resource from the web can reached a consistent explaining, then ambiguity resolved.

## 3 Understanding of Collective Wisdom Based on Semantic Web

Form above introduce of semantic web, it is known that the discussion content, resource description, services invoking in HWME can be descript extensively by semantic web, thus, a great access to the each part in HWME can be reached. Because of the obvious advantages [11-13], semantic web is the naturally chosen for HWME.

### 1) Extensibility

The research objects in HWME, often open complex giant system, involves many scientific fields and system resources; and many experts from various fields bring out new conceptions and academic thoughts; thus, the way of knowledge expressing in HWME should be opening and extendible, and can help human knowledge evolution. [14] In RDF, RDFS, user defined ontologies and reasoning rules, helping to extend academic space in discussion, just satisfies the requirement of HWME.

### 2) Self-explainable

RDF, RDFS are based on XML, its semantic is self-explainable[9]. So, both the communicators in discussion can understand each other by messages itself.

### 3) Semantic definiteness

The basic theory of RDF, RDFS, OWL is ontology, which is refine result of common knowledge in some domains, the object with same intension and extension is unique in semantic web, and expressed by unique URI. If some experts have different description about the same object, they should use union algorithm to get one URI, and then resolve the ambiguity.

### 4) Logic reasoning

RDF, RDFS, OWL, involves artificial intelligence, especially first order predicate logic into semantic web, which help HWME is capable of logic reasoning, so, the parts in HWME can get help to make right judgment.

However, how to express the knowledge, resource, attributes and relationships in HWME by semantic web?

## ①process of understanding collective wisdom based on semantic web

In HWME, at the beginning of discussion, in common sense, the chairman has primitive knowledge about the content of the discussion, so he can make the first original ontology model, most important partly to create the topic relative RDF Schema, and try to express the relative domain knowledge. Then, some experts invited by chairman give their suggestions to assure this topic's scale, content, goal and estimation, the first model maybe needs verifying, improving or pushing over, but after affords given, the expert group can be set up, and the chairman and newly expert group can get one more advanced semantic web model, in which, the fully vocabulary, schema and initial construction of resources and attributes are put forward. The parts in the discussion can express their own opinion based on these primitive model copies, experts group can examine, fix and improve this model to get their own version. These versions will be saved individually, every change, causes and objects will be saved completely, which are necessary condition in effective mutual communication model [15], also are important bases in the coming consensus analyses and estimation. With the deepening of discussion, if the increasing consistency can be reached, the versions of the model can be consisted gradually, finally, a consistent semantic web model can be achieved, in which, all the statements can express one only truth; but in another situation, the conclusion should be got by consensus analyses and estimation.

## ②semantic web knowledge expression of role speaking

In discussion, the role speaking is the most important part, so do the semantic web knowledge expression of role speaking. Role speaking has two kinds of forms, one is

speaking in languages, and the other is using dialog tools, such as msn. In HWME, both the two forms, should be expressed by semantic web.

To the first form, the steps as follows:

- a) Put the content talking about into files by voice input or manual input;
- b) According to the current semantic web model version, find and match the content to the schemas, create the new attributes or change its values;
- c) To the content which can not be expressed automatically, using RDF, RDFS, OWL modeling tools to add into current model version, if new word created, update the RDF schema and reasoning machine synchronously.

To the second form, we can use the dialog tools which support semantic web modeling function. Parts in HWME can communication with each other and express their opinions in a visual way in these tools.

## 4 Achieving Consensus of Discussion

The goal of HWME is to achieve a more intensive consensus of topics, however, how to estimate the consistent level in communication, which is the key to end the discussion.

In some situation, experts is the leader in whole procedure, so naturally, experts can judge when to end the discussion by their knowledge and experiences. Expert group is the soul and spirit in HWME, any conclusion or no conclusion can be drawn should be confirmed by experts group.

When expert group achieve a consistent conclusion about topic in discussion, they fix one semantic web model version to get a perfect one, and output the final solution by this version; when expert group think it is not the right time to draw any conclusion, all the versions should be reserved to wait more investigation; when expert group can not access to perfect solution, but matters need a quite answer, so they can get a better one by links construction analysis method.

Links construction analysis method presented by CUI [6], who think the links construction build through mutual communication between experts. In process of mutual communication, experts give their opinions and response to the others, thus a links construction direction graph build according to these responding relations. Each speech is the node in this graph, and each node has two attributes, one is quality attribute, named Authority, and the other is responding attribute, named Hub. Each responding or responded speech is directed link arc, which is an enhancing relationship between the nodes because there is positive feedback effect between speakers in discussion. So, there are two iterative formulas to compute Authority values and Hub values:

$$a(pt) := \sum_{qt \rightarrow pt} h(qt) \quad (2)$$

$$h(pt) := \sum_{pt \rightarrow qt} a(qt) \quad (3)$$

In which,  $h(pt)$  is every speech's Hub values,  $a(pt)$  is Authority values, all initial values equal 1.  $pt \rightarrow qt$  means speech  $pt$  responding to speech  $qt$ .

After compute nodes' two attributes values, and sort the values, find out the nodes with the greater values, which are opinions or suggestions focused on in the discussion. This method is simple, obvious and robust, but it only takes the responding relationship between nodes into account, and not thinks about the characters of responses, such as agreement, partial agreement, rejection, partial rejection, query, asking, and etc. Moreover, this method can find out the major club about topic, but some issues which have no doubt will be ignored because of their lower Hub values.

In this paper, HWME based on semantic web is presented, in which, all resources and speeches are expressed by semantic web, and enough information supplied for whole process and the conclusion; and with the capability of logic reasoning of semantic web, every role in discussion can validate his opinion by semantic web model. So, in HWME based on semantic web, the final consensus of discussion can be reached by these steps:

①Each role verifies his own semantic web model version, assuring that model can express his opinion about the topics in discussion completely, exactly and clearly.

②Each role commits his own semantic web model version, and union all versions into a gross one, then make consistence verification. Due to the same basic RDF Schema, the gross version have the same definition of resources and attributes, and the consistent part of result can be reserved completely, but with the differences in roles, some conflicts occur in gross version.

③Analyses the conflicts one by one, find out the relative resources, attributes and their relationships, and index on resource, exam the records of speeches in which the resources and attributes have responses. These speeches and responses also build a links construction graph, so the Authority values and Hub values about relative nodes in this graph can be figured out. In this paper, with the semantic and RDF schema, we can change equ.(2), equ.(3) into:

$$a(pt) := \sum_{qt \rightarrow pt} h(qt) - \sum_{qt \rightarrow pt} \bar{h}(qt) \quad (4)$$

$$h(pt) := \sum_{pt \rightarrow qt} a(qt) - \sum_{pt \rightarrow qt} \bar{a}(qt) \quad (5)$$

In which,  $h(pt)$  means the agreement Hub values, otherwise,  $\bar{h}(pt)$  means the disagreement Hub values, and due to resource is just one node, and then the relationship of partial agreement and partial rejection do not exist.  $a(pt)$  means the speech Authority values, while  $\bar{a}(pt)$  means inverse relationship Authority values, all initial attributes values equal 1.  $pt \rightarrow qt$  means speech  $pt$  responding to speech  $qt$ .

④Iteratively compute equ.(4) and equ.(5), if final result  $h(pt) > 0$ , then the relationship between the node and its attributes are right, which should be reserved in the gross version, the conflicted links of resources and attributes should be removed. Otherwise, keep the conflicted ones, remove this one.

⑤Go loop with all conflicts in gross version until verification finish.

⑥According to the final gross version, output the statements set, achieving the final conclusion.

## 5 Conclusion

In HWME, how to realize the mutual communication and understanding between man and man, man and equipment, and man and system is the key to stimulate collective wisdom. This paper presents a semantic network as a bridge for communication between those different parts. This method expresses knowledge by resources, attributes and semantic relationships, and resolving the ambiguity between nodes by union algorithm, moreover, the capability of logic reasoning in HWME can help the roles in discussion to make right judgment. In the practice of Ge-La pipeline leaking detective HWME, it is proved that this method is effective.

**Acknowledgments.** Founds Of Nature Science in ChongQing (ID: 2005BB6215).

## References

1. Qian, X.-S., Yu, J.-Y., Dai, R.-W.: A New Discipline Of Science: The Study Of Open Complex Giant System And Its Methodology. *Chinese Journal of Nature*, 3–10 (1990) (in Chinese)
2. Dai, R.-W.: Complex Giant System Science: Science in 21st. *Chinese Journal of Nature*, 187–192 (1997) (in Chinese)
3. Dai, R.-W., Li, Y.-D.: Researches on Hall for Workshop of Metasynthetic Engineering and System Complexity. *Complex Systems and Complexity Science*, 1–24 (2004) (in Chinese)
4. Cui, X., Dai, R.-W.: Human-centered Artificial Society of Cyberspace of Meta-synthetic Engineering for Workshop I. *Complex Systems and Complexity Science*, pp. 1–8 (2006) (in Chinese)
5. Cui, X., Dai, R.-W.: Human-centered Artificial Society of Cyberspace of Meta-synthetic Engineering for Workshop II. *Complex Systems and Complexity Science*, pp. 21–27 (2006) (in Chinese)
6. Cui, X., Dai, R.-W., Li, Y.-D.: Research on Happening of Collective Wisdom in Hall for Workshop of Metasynthetic Engineering. *Journal of System Simulation*, 24–28 (2003) (in Chinese)
7. Tan, J.-F., Zhang, P.-Z., Huang, L.-N.: A Group Argumentation Information-Structuring Model in Hall for Workshop of Metasynthetic Engineering. *Systems Engineering-Theory & Practice*, 34–38 (January 2005) (in Chinese)
8. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American*, 284–289 (2001)
9. Miller, E.: An Introduction to the Resource Description Framework. *D-Lib Magazine*, 23–26 (1998)
10. Li, J., Ding, Y.: Semantic Web, Semantic Grid and Semantic Network. *Computer and Modernization*, 38–41 (2007) (in Chinese)

11. De Roure, D., Jennings, N.R., Shadbolt, N.R.: The semantic grid: Past, Present and Future. *Proceeding of the IEEE*, 93–96 (2005)
12. Tian, C.-H.: Research on Semantic Web in China. *Journal of The China Society for Scientific and Technical Information*, 243–249 (2005) (in Chinese)
13. Song, W., Zhang, M.: A First Top Towards The Semantic Web. Higher Education Press (2004) (in Chinese)
14. Wei, H., Pan, Y.-H.: From Knowledge Expression to Expression: the Advancement in Artificial Intelligence Research. *Journal of Computer Research and Development*, 819–825 (2000) (in Chinese)
15. Janssen, T., Sage, A.P.: A support system for multiple perspectives knowledge management and conflict resolution. *International Journal of Technology Management*, 19–22 (2000)



# A General Result on the Convergence of Online BP Learning

Rui Zhang<sup>1</sup> and Zuo-Zhi Liu<sup>2</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798

<sup>2</sup> Department of Mathematics, Northwest University, Xi'an, 710069, China  
zh0010ui@e.ntu.edu.sg, egbhuang@ntu.edu.sg

**Abstract.** The backpropagation (BP) algorithm has been widely applied for training feedforward neural networks in scientific research and engineering. The success of the application, however, relies upon the convergence of the learning procedure. In this paper, we further analyze the convergence issue and provide a general result on convergence of the online BP learning algorithm. We prove that under mild conditions, the weight sequence defined by the learning procedure converges to a fixed value at which the error function attains its minimum.

**Keywords:** Online BP learning, feedforward neural networks, convergence analysis.

## 1 Introduction

The backpropagation (BP) algorithm is widely used for training feedforward neural networks with which the network weights are updated according to the gradient method [1], [4]. In general, there are two ways to implement the gradient method: either the batch scheme or the online scheme. The batch scheme corresponds to the standard gradient iteration procedure which updates the network weights after all the training samples are processed. Differently, with the online scheme, all the training samples are sequentially (one-by-one) presented to the learning system and only one sample is leaned at each time. The learned sample may be randomly or circularly selected from the given training samples, but should keep periodic in the training set. The online BP learning is generally preferred over batch learning in some applications especially when the given training examples are huge. Therefore, it is worth mentioning the convergence of such online BP learning algorithm.

The deterministic convergence of online BP learning has been analyzed in [2], [3], [5]-[12]. The neural networks discussed in [3], [5], [6], [8], [9], [11] are of two-layer (say, without hidden elements), which are of special form, and hence, have limited capabilities in applications. Therefore, it is meaningful and necessary to study the convergence of the online BP learning algorithm in the case for single-hidden-layer feedforward neural networks. Wu et al. [2] [7] prove that the gradient sequence of the error functions converges to zero with some impractical constraints on setting of the step-size, which cannot be verified before the BP learning starts. Furthermore, Xu *et al.* [10]

improved the result obtained in [7] through proving two fundamental theorems on convergence of the online BP learning procedure. One theorem claims that under mild conditions, the gradient sequence of the error function will converge to zero, and another theorem concludes that the weight sequence defined by the procedure converges to a fixed value at which the error function attains its minimum.

In this paper, we present a generalized deterministic analysis on convergence of online BP learning for the single-hidden-layer feedforward neural networks. We will prove that the weight sequence generated by the online BP learning algorithm converges to a fixed value at which the error function attains its minimum in much weaker conditions than that assumed in [10].

The remainder of the paper is organized as follows. In the next section, we review the online BP learning algorithm. We present the main results and the rigorous proof in Section 3 with a series of necessary lemmas. We then conclude the paper in Section 4.

## 2 Brief of Online BP Learning Algorithm

In this paper, we consider single-hidden-layer feedforward neural networks with  $p$  inputs,  $n$  hidden neurons and 1 output neuron. The activation functions used in the hidden and output neurons are all the same, a continuously differentiable function, denoted henceforth by  $g : R \rightarrow R$ . Denote by  $V = (v_{ij})_{n \times p} = (v_1, v_2, \dots, v_n)^T$  ( $v_i = (v_{i1}, v_{i2}, \dots, v_{ip}), i = 1, \dots, n$ ) the weight matrix connecting the input and hidden layers, and by  $W = (w_1, w_2, \dots, w_n)^T$  the weight vector connecting the hidden and output layers. Given the training sample set  $D_J = \{(\xi^j, o^j) : \xi^j \in R^p, o^j \in R\}_{j=0}^{J-1}$ , we define the error function of the neural network is

$$E(W, V) = \sum_{j=0}^{J-1} E_j(W, V) = \sum_{j=0}^{J-1} \frac{1}{2} (o^j - g(W^T G(V \xi^j)))^2.$$

where  $E_j(W, V)$  is called the individual error function and  $G(x) = (g(x_1), g(x_2), \dots, g(x_n))^T, \forall x \in R^n$ . Denote by  $\nabla E_j(W, V)$  and  $\nabla E(W, V)$  the gradients of the individual error function  $E_j(W, V)$  and the error function  $E(W, V)$  with respect to  $(W, V)$  respectively, then we have  $\nabla E(W, V) = \sum_{j=0}^{J-1} \nabla E_j(W, V)$  (here  $(W, V)$  represents the whole weight matrix  $[W, V]$ ). It has been proved that both  $\nabla E_j(W, V)$  and  $\nabla E(W, V)$  are bounded and Lipschitz continuous (Lemma 2, [12]). Throughout this paper, we use the notation  $\Omega(E) = \{(W, V) : \nabla E(W, V) = 0\}$  to denote the set of the stationary points of the error function  $E(W, V)$ .

With the above notations and equations, the online BP learning algorithm can be formulated as the following iteration procedure [10]:

$$\begin{aligned} & (W^{mJ+j+1}, V^{mJ+j+1}) \\ &= (W^{mJ+j}, V^{mJ+j}) - \eta_m \nabla E_j(W^{mJ+j}, V^{mJ+j}) \\ & \quad j = 0, 1, \dots, J-1; m = 0, 1, \dots \end{aligned} \tag{1}$$

In procedure (1), the weight sequence  $\{(W^{mJ+j}, V^{mJ+j})\}$  is updated incrementally at each cycle  $(m = 0, 1, \dots)$  through  $J(j = 0, 1, \dots, J-1)$  iterations. In each iteration, the weights are updated along the direction of negative gradient of an individual error function associated with a single training sample. In this paper, we only consider the case of choosing the training samples in a fixed order, that is, from  $\xi^0$  to  $\xi^{J-1}$  one by one sequentially.  $\eta_m$  is the step-size of the procedure, whose value is updated after all the training samples are fed and satisfies

$$\eta_m > 0, \quad \sum_{m=0}^{\infty} \eta_m = \infty \quad \text{and} \quad \sum_{m=0}^{\infty} \eta_m^2 < \infty.$$

### 3 Main Results

In this section, we will present a general result on convergence of the weight sequence  $\{(W^{mJ+j}, V^{mJ+j})\}$  generated by the online BP learning algorithm (1) under the following assumption:

(S1).  $|g(t)|, |g'(t)|$  are uniformly bounded for  $t \in R$  and  $g'(t)$  is Lipschitz continuous, that is, there is a positive constant  $L$  such that

$$|g'(t) - g'(\bar{t})| \leq L |t - \bar{t}|, \quad \forall t, \bar{t} \in R;$$

(S2). The weight sequence  $\{\| (W^k, V^k) \|: k = 1, 2, \dots\}$  is upper bounded.

(S3). The set  $\Omega(E)$  has no connected subset containing more than one point.

As mentioned in the proceeding section, with the online BP training procedure (1), the weight sequence  $\{(W^{mJ+j}, V^{mJ+j})\}$  is updated incrementally at each cycle  $(m = 0, 1, \dots)$  through  $J(j = 0, 1, \dots, J-1)$  iterations. In the subsequent convergence analysis, we first clarify the convergence results for the sequence  $\{(W^{mj}, V^{mj})\}$ , which is composed of the elements generated at the beginning of each cycle. Then we extend the obtained results to the general case of the whole weight sequence  $\{(W^{mJ+j}, V^{mJ+j})\}$ .

**Lemma 1.** (Lemma 3, [12]) Let  $\{E(W^{mj}, V^{mj})\}$  be the error sequence generated by the online BP learning algorithm (1). Under assumptions (S1) and (S2), there exists a constant  $C$  such that

$$E(W^{(m+1)J}, V^{(m+1)J}) - E(W^{mJ}, V^{mJ}) \leq -\eta_m \|\nabla E(W^{mJ}, V^{mJ})\|^2 + C\eta_m^2 \tag{2}$$

**Lemma 2.** (Theorem 2, [12]) Let  $\{E(W^{mJ}, V^{mJ})\}$  be the sequence of the error functions generated by the online BP learning algorithm (1). Under assumptions (S1) and (S2), there exists  $(W^*, V^*) \in \Omega(E)$  such that

$$\lim_{n \rightarrow \infty} E(W^{mJ}, V^{mJ}) = E(W^*, V^*).$$

**Lemma 3.** (Lemma B-2, [10]) Under the assumptions (S1) and (S2), every limit point of the weight sequence  $\{(W^{mJ}, V^{mJ})\}$  is the stationary point of the error function  $E(W, V)$ .

Denote by  $H$  the set of the limit points of the weight sequence  $\{(W^{mJ}, V^{mJ})\}$ , then from Lemma 3, there obviously has  $H \in \Omega(E)$ .

**Theorem 1.** Under the assumptions (S1), (S2) and (S3), the set  $H$  is closed, compact and connected.

*Proof:* It is obvious that  $H$  is closed and compact according to the definition of  $H$  and Lemma 2.

We next prove the connectivity of  $H$  by the contradictory method. Assume that  $H$  is not connected, then according to the compactness of  $H$ , we know that  $H$  can be expressed as the union of two disjoint nonempty compact subsets  $H_1$  and  $H_2$ , that is,  $H = H_1 \cup H_2$  with  $H_1 \neq \emptyset$ ,  $H_2 \neq \emptyset$  and  $H_1 \cap H_2 = \emptyset$ . Suppose that  $\rho(H_1, H_2) = \varepsilon$  where

$$\rho(H_1, H_2) = \inf\{\rho((W_1, V_1), (W_2, V_2)) : (W_1, V_1) \in H_1, (W_2, V_2) \in H_2\}$$
 and let

$$G_1 = \{(W, V) : \rho((W, V), H_1) < \frac{\varepsilon}{3}\},$$

$$G_2 = \{(W, V) : \rho((W, V), H_1) < \frac{2\varepsilon}{3}\},$$

$$G_3 = \{(W, V) : \rho((W, V), H_1) < \varepsilon\},$$

then  $G_3 \cap H_2 \neq \emptyset$ . We then conclude that there are finite elements of  $\{(W^{mJ}, V^{mJ})\}$  in  $\overline{G_2} - G_1$ , since otherwise, there must be  $\overline{G_2} \cap H_2 \neq \emptyset$ , which obviously contradicts to  $G_3 \cap H_2 \neq \emptyset$ . Here,  $\overline{G_2}$  denotes the closure of  $G_2$ . Therefore, there exists a positive number  $M_1$  such that  $(W^{mJ}, V^{mJ}) \notin \overline{G_2} - G_1$  for all  $m$  with  $m \geq M_1$ .

Let  $M = \max\{M_1, M_2\}$ . Then the sequence  $\{(W^{mJ}, V^{mJ})\}$  has the following property:

(\*) If  $(W^{mJ}, V^{mJ}) \in G_1$ , then  $(W^{(m+1)J}, V^{(m+1)J}) \in G_1$ , This is because if  $(W^{mJ}, V^{mJ}) \in G_1$  with  $m \geq M$ , that is

$$\rho((W^{mJ}, V^{mJ}), H_1) < \frac{\varepsilon}{3},$$

then according to (2), we have

$$(W^{(m+1)J}, V^{(m+1)J}) \leq (W^{mJ}, V^{mJ}) + z_m$$

We then obtain

$$\rho((W^{(m+1)J}, V^{(m+1)J}), H_1) < \frac{2\mathcal{E}}{3},$$

that is,  $(W^{(m+1)J}, V^{(m+1)J}) \in G_1$ . However, according to the conclusion shown above that there is no element of  $\{(W^{mJ}, V^{mJ})\}$  in  $\overline{G_2} - G_1$  for all  $m$  with  $m \geq M$ , we then conclude that there must be  $(W^{(m+1)J}, V^{(m+1)J}) \in G_1$ .

For any  $(\overline{W}, \overline{V}) \in H_1$ , there exists a positive number  $\overline{m}$  such that  $\overline{m} \geq M$  and  $(W^{\overline{m}J}, V^{\overline{m}J}) \in G_1$ . Therefore, by the property (\*), we know that  $(W^{(\overline{m}+1)J}, V^{(\overline{m}+1)J}) \in G_1$ .

Hence, for all  $m \geq \overline{m}$ , there is

$$(W^{mJ}, V^{mJ}) \in G_1.$$

This is obviously contradict to the facts that  $H_2 \neq \emptyset$  and every element of  $H_2$  is the limit point of  $\{(W^{mJ}, V^{mJ})\}$ .

This contradiction verifies the connectivity of  $H$ .

**Theorem 2.** Let  $\{(W^{mJ}, V^{mJ})\}$  be the weight sequence generated by the online BP learning algorithm (1). Under assumptions (S1), (S2) and (S3), there exists  $(W^*, V^*) \in \Omega(E)$  such that

$$\lim_{n \rightarrow \infty} (W^{mJ}, V^{mJ}) = (W^*, V^*).$$

*Proof:* According to Lemma 3, we know that  $H \in \Omega(E)$ . By assumption (S3), if  $\Omega(E)$  has no connected subset containing more than one point, then  $H$  has no connected subset containing more than one point. Since  $H$  is closed, compact and connected from Theorem 1, we can immediately conclude that  $H$  can only be a single-point set, that is to say, the limit points of the sequence  $\{(W^{mJ}, V^{mJ})\}$  is unique. Denote it by  $(W^*, V^*)$ , we then have

$$\lim_{n \rightarrow \infty} (W^{mJ}, V^{mJ}) = (W^*, V^*).$$

In view of the continuity of the error function  $E$  and Lemma 2, we know that  $\nabla E(W^*, V^*) = 0$ , that is to say,  $(W^*, V^*)$  is a stationary point of  $E(W, V)$ . The proof of Theorem 2 is thus completed.

Note that in the case where  $E(W, V)$  has at most countable infinite number of stationary points (say, assumption (B1) in [10]), it is obvious that the set of the stationary points  $\Omega(E)$  is either finite or infinitely countable. Under either of above situations,  $\Omega(E)$  has no connected subset containing more than one point. Therefore, condition (B1) in [10] is just a special case of (S3), and hence, Theorem B-1 in [10] can be viewed as a corollary of Theorem 2 in this paper. To be more precise, we have

**Corollary 1.** Let  $\{(W^{m_j}, V^{m_j})\}$  be the weight sequence generated by the online BP learning algorithm (1). Under assumptions (S1) and (S2), if  $E(W, V)$  has at most countable infinite number of stationary points, then there exists  $(W^*, V^*) \in \Omega(E)$  such that

$$\lim_{n \rightarrow \infty} (W^{m_j}, V^{m_j}) = (W^*, V^*).$$

The following corollary shows that the convergence of the weight sequence can also hold under the case where  $\Omega(E)$  is infinitely uncountable.

**Corollary 2.** Let  $\{(W^{m_j}, V^{m_j})\}$  be the weight sequence generated by the online BP learning algorithm (1). Under assumptions (S1) and (S2), if  $\Omega(E)$  is a Cantor set, then there exists  $(W^*, V^*) \in \Omega(E)$  such that

$$\lim_{n \rightarrow \infty} (W^{m_j}, V^{m_j}) = (W^*, V^*).$$

Corollary 2 can be directly proved since a Cantor set is infinitely uncountable and consistent with assumption (S3).

## 4 Conclusions

In this paper we have analyzed the deterministic convergence of online BP learning for single-hidden-layer feedforward neural networks. A general convergence theorem has been proven, claiming that the weight sequence  $\{(W^{m_j+j}, V^{m_j+j})\}$  generated by online learning converges to a fixed value at which the error function attains its minimum under mild conditions. The obtained convergence theorem generalizes those existing analysis (particularly, the results obtained recently by Xu *et al.* [10]) in the sense that our analysis relaxed condition (B1) assumed in [10] into the more general and weaker condition (S3).

**Acknowledgments.** This work was supported by the National Natural Science Foundation under Grant 61075050, China.

## References

1. Haykin, S.: Neural networks: a comprehensive foundation. Macmillan College Publishing Company, New York (1994)
2. Li, Z., Wu, W., Feng, G., Lu, H.: Convergence of an Online Gradient Method for BP Neural Networks with Stochastic Inputs. In: Wang, L., Chen, K., Ong, Y.S. (eds.) ICNC 2005. LNCS, vol. 3610, pp. 720–729. Springer, Heidelberg (2005)
3. Li, Z.-X., Wu, W., Tian, Y.-L.: Convergence of an online gradient method for FNN with stochastic inputs. Journal of Computational and Applied Mathematics 163(1), 165–176 (2004)
4. Looney, C.G.: Pattern recognition using neural networks: theory and algorithms for engineers and scientists. Oxford University Press, New York (1997)

5. Luo, Z.-Q., Tseng, P.: Analysis of an approximate gradient projection method with application to the backpropagation algorithm. *Optimization Methods & Software* 4(2), 85–101 (1994)
6. Wu, W., Feng, G.-R., Li, X.: Training multilayer perceptions via minimization of sum of ridge functions. *Advances in Computational Mathematics* 17(4), 331–347 (2002)
7. Wu, W., Feng, G.-R., Li, Z.-X., Xu, Y.-S.: Deterministic convergence of an online gradient method for BP neural networks. *IEEE Transactions on Neural Networks* 16(3), 533–540 (2005)
8. Wu, W., Shao, Z.-Q.: Convergence of an online gradient methods for continuous perceptrons with linearly separable training patterns. *Applied Mathematics Letters* 16(7), 999–1002 (2003)
9. Wu, W., Xu, Y.-S.: Deterministic convergence of an online gradient method for neural networks. *Journal of Computational and Applied Mathematics* 144(1) (2002)
10. Xu, Z.-B., Zhang, R., Jing, W.-F.: When does online BP training converge? *IEEE Transactions on Neural Networks* 20(10), 1529–1539 (2009)
11. Zhang, N.-M., Wu, W., Zheng, G.-F.: Convergence of gradient method with momentum for two-layer feedforward neural networks. *IEEE Transactions on Neural Networks* 17(2), 522–525 (2006)
12. Zhang, R., Yang, L., Wang, W.: A new Convergence Property of Online BP Learning. In: *Proceeding of International Conference on Information Science and Technology (ICIST 2011)*, Nan jing, China, March 26-28, pp. 231–237 (2011)

# Research and Application of Data Synchronization Based on Group Communication Technology

Li-xin Li and Jue Wang

Information Technology College, Zhengzhou, Henan, China

**Abstract.** To solve the problem of data consistency in distributed database system, this paper presents a global transaction-oriented synchronization model based on the features of total order service in group Communication. This model combines the transaction on each node with the technology of Atomic Broadcast and guarantees the data consistency of the system strictly. Results show that this model reduces the traffic of synchronization and improve the efficiency of eager replication.

**Keywords:** data synchronization, Group Communication, Atomic Broadcast, global transaction.

## 1 Introduction

Distributed database system [1] (Distributed Database System, DDBS) includes distributed database (DDB) and distributed database management system (DDBMS). Among them, the distributed database [2] is physically located in different nodes of a computer network, but logically they belong to the same system of data collection. In other words, the distributed database is a structured collection of data belonging to the same logical system, and computer networks are physically distributed across all nodes.

In a distributed database system, in order to improve system availability, enhance system fault tolerance [1,2], commonly used data synchronization technology. If two or more distributed database servers are stored the same data and provide equivalent service, customers can choose from a failed node to another node in the normal operation, a node failure will not lead to paralysis of the system, which strengthen the availability of the service, and also improve the system's fault tolerance. Generally, we used technology to achieve data synchronization between distributed databases. Since the data synchronization in a distributed database system on multiple nodes, so each node to maintain data consistency is the core technology research database synchronization problems.

## 2 Data Synchronization

### 2.1 Data Synchronization

In the field of data integration and information sharing of a distributed database system, database synchronization is a core technology. Database synchronization



refers to two or more different databases to exchange data between nodes, so that one of the database data changes occurred in other nodes in the database will be shown accordingly. One-way database synchronization can also be a two-way [3]. One-way sync, also known as master-slave synchronization, the database will sync into the main database and from the database. The data of the master database is based data, only the main changes will be synchronized to the slave database from the master database. Two-way synchronization is the database system, each node of the relationship as based from, any database changes, the other nodes corresponding synchronization.

## 2.2 Synchronous Replication and Asynchronous Replication

Data synchronization protocol is used to solve the data consistency of each node in the database system. Based on the updated time synchronization protocol can be divided into synchronous replication and asynchronous replication [1].

Synchronous replication is a real-time data access and synchronization of distributed real-time update technique, also known as real-time replication. Synchronous replication is the copying of any node in the updating and dissemination of data to carry out a transaction, the transaction was only really commit after the update operation completed on all nodes, otherwise the entire transaction is rolled back. This mechanism ensures that any time a copy of each node on the strict data consistency, but the increased consumption of the system, and the longer response because of maintaining strict data consistency.

Asynchronous replication is also known as storage-forwarding replication, the update operation is only a copy spread to other nodes after the transaction is committed. Copy data between nodes can be temporarily out of sync, but in the end will ensure that copies of the data on all nodes is consistent. Asynchronous replication has the advantage of reducing the consumption of system communication and shortening the response time. However, the delay in the submission of transaction causes data being inconsistent on the nodes of the system, sometimes resulting data conflicts.

At present, commercial database products typically focus on system efficiency in practical application, they tend to use asynchronous replication protocol, achieve the weak consistency at the expense of sacrificing data consistency. Thus, an efficient Synchronous replication is needed to meet the demand of strict data consistency and real-time in distributed systems. In this paper, we consider characteristics of synchronous replication and asynchronous replication, group communication technology based on atomic broadcast mechanism, combined with a global transaction for data synchronization strategy, while reducing the traffic system to ensure the consistency of data between nodes.

## 3 System Model and Related Technologies

### 3.1 System Model

In this paper, the distributed databases system is divided into multi-domain and multi-synchronous architecture, which contains a number of regions, each region has a

number of databases. Within the region, according to the relationship between master and slave nodes into the central node and the slave nodes; the center node that needs to receive the data synchronization message is a member of group communication systems. Group members synchronize data with each other between the regions via group communication support services. Structure of the system is shown in Figure 1.

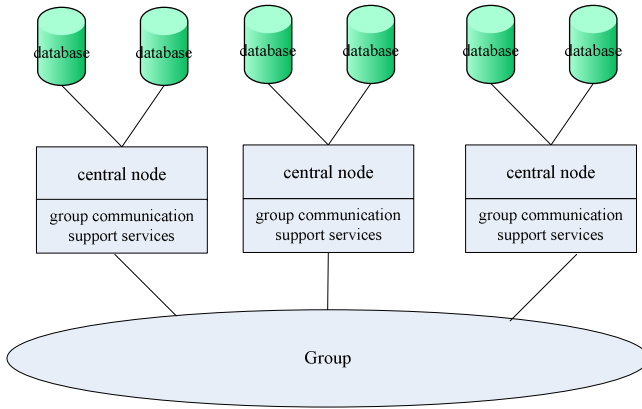


Fig. 1. Group communication system structure

### 3.2 Group Communication and Atomic Broadcast

#### 3.2.1 The Group Communication Mechanism

Group communication mechanism is a distributed database system in a number of relevant nodes as a group; a member of the group node send or broadcast data synchronization information to all members of the group; the group communication support services on each member node received information, processed and forwarded to the local database system implementation. Group communication is an effective safeguard to share information. Distributed database system for data synchronization, how to ensure the arrival of synchronization of information to other nodes and the consistent order of information is a major consideration, we can solve the information dissemination and sequencing problems in synchronous replication based on the use of atomic broadcast mechanism and message total ordering technology of group communication mechanism.

#### 3.2.2 Atomic Broadcast

On the order of information dissemination, group communication systems provide several different types of forwarding the order [4]:

(1) Basic service. When a node receives a message from the other node, the information is immediately delivered. Since the order of the information on each node may be different, basic service therefore can not guarantee the order to receive information.

(2) First in first out service. This is also known as FIFO order, the information in accordance with “first in first out” principle for transmission and delivery.

(3) Causal order service. Also known as the causal sequence, if the causal priority message  $m$  in  $m'$ , then  $m$  are first in  $m'$  is delivered on all member nodes in the group communication system.

(4) Total order service. This is also known as Atomic Broadcast, the information was delivered in the same order on all nodes of the group communication system. In other words, the system node  $N_1$  and  $N_2$ , receive message  $m$  and  $m'$ , in 和  $N_1$  and  $N_2$ , or  $m$  the first delivery after the delivery of  $m'$ ; or the first delivery of  $m'$ , after the delivery of  $m$ . The order of Information delivery in all members of the nodes in group communication systems is exactly the same.

The communication model in the group communication system is based on atomic broadcast technology. Atomic broadcast technology mainly based on the sequence information to define, to ensure the submission to all nodes in the communication system is delivered on in the same order. The main features of atomic broadcast [5,6] are:

(1) All-or-nothing feature. If a group broadcast a message, or all members of the group receive the message correctly, or all not received. Will not allow some members of the group received some did not receive the message.

(2) The message sequencing features. If a member of the group sequential delivery of messages  $m$  and  $m'$ , then  $m$  and  $m'$  will be delivered in the same order on all the other nodes of the group members.

Using the atomic broadcast group communication protocol for messaging in a distributed database system can bring the following benefits:

(1) Reduce the system communication response time. Atomic Broadcast communication eliminates a clear response on the information at the application level, which can reduce the execution due to communication latency.

(2) Ensure data consistency. All-or-nothing feature of Atomic broadcast ensures the data consistency of nodes within the group.

(3) The elimination of global synchronization. Because all sites are appropriate in the progress of local arrangements, which can deal with the global deadlock does not require any synchronization.

## 4 Global Transaction-Oriented Data Synchronization Model

In this paper, we present a global transaction-oriented synchronization model based on the features of total order service in group communication. This model combines group communication mechanism and transaction mechanism, reduces consumption of the system and improves data synchronization performance.

### 4.1 The Structure of Communication Support Services

Group communication support services intercepts the data between the user and the database exchange, and synchronizes the submission of transaction request to all nodes of the group to executive simultaneously, to achieve transparent database synchronization.

As shown in Figure 2, group communication support service consists of four parts: transaction processing module, data synchronization protocol, database access engine and communications management module. Transaction processing module is divided into transaction intercept module and transaction analysis module.

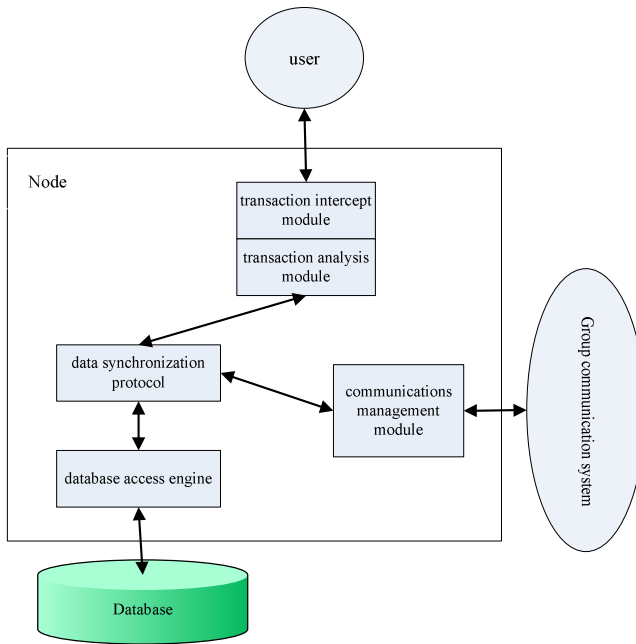
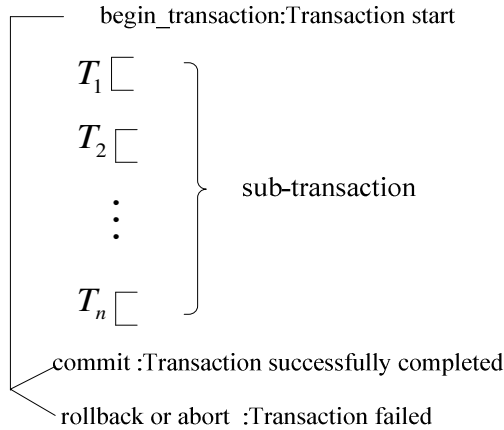


Fig. 2. Structure of group communication support services

### 4.2 Global Transaction-Oriented Data Synchronization Strategy

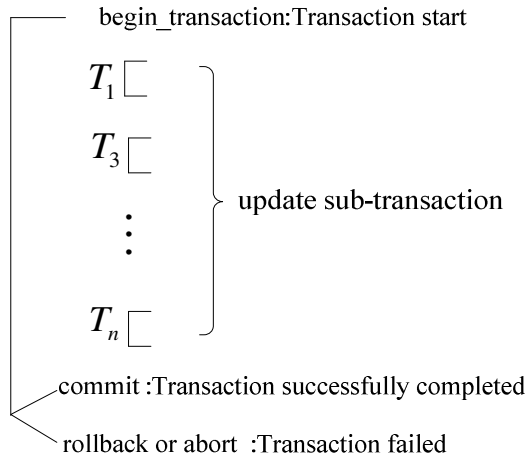
There are a few concepts before explaining the synchronization model:

(1) Local transaction: this involves only a single node in the local execution of transactions for one or more operations. The transaction begins with 'begin\_transaction' symbol; when the transaction is successfully completed, it ends with commit symbol, otherwise ends with 'rollback' or 'abort' symbol. Transaction structure is shown in Figure 3.



**Fig. 3.** Structure of local transaction

(2) Global transactions: After processing the local transactions, system will ultimately determine the global order of these transactions through the atomic broadcast. Transaction structure is shown in Figure 4; sub-transaction  $T_2$  is read-only transaction, submit it and broadcast the rest of update sub-transactions. We get the order of  $T_1, T_3 \dots T_n$  after broadcasting the sub-transactions. This transaction is defined as a global transaction.



**Fig. 4.** The structure of global transaction

In global transaction-oriented data synchronization strategy, the main idea is to allow local first concurrent transactions executed by the DBMS to ensure that the transaction is locally serializable, and record the process of implementation of local

transactions; when receive the end symbol of the local transaction, make the order of a local implementation in accordance with the sequence that has received submissions; then the ordered transaction will be broadcast by the group communication system. So we can so we can ensure that the global order of execution and local consistency. When the broadcast transaction and the local transaction conflict, if the local transaction in the local implementation phase, you want to abort the local transaction. If the local transaction has been sent, the transaction has been fully accessible and to determine the global order. After the transaction undo and redo in accordance with its global order, in order to reduce the rate of withdrawal transactions. When received the news of the success of the implementation from all nodes, the global transaction is committed, otherwise rolled back the transaction.

Global transaction-oriented data synchronization algorithm is as follows:

Suppose node  $N_1$  and  $N_2$  are members of group Communication systems, transaction  $T_1$  is performed on  $N_1$ , transaction  $T_2$  is performed on  $N_2$ .

### **Step 1: The local execution and recording of transaction**

Step 1.1 The user submit  $T_1$  on  $N_1$ , then the system to determine whether  $T_1$  conflict with the global transaction, abort  $T_1$  if there is conflict and close the transaction;

Step 1.2 Record  $T_1$  until the transaction processing module receives the symbol that submitted by  $T_1$ .if the rollback symbol, close the transaction;

Step 1.3 system determines a local order according to the received commit symbol, and organizes all update transaction into global transaction  $T_1'$ .

### **Step 2: Broadcasting**

Step 2.1 If  $T_1$  is Read-only transaction, submit  $T_1$  on  $N_1$ , end.

Step 2.2 Broadcast  $T_1'$  to all nodes of the group by the group communication system, and then identify  $T_1$  with 'executed' state.

### **Step 3: Implementation of global transaction**

Step 3.1 Node  $N_2$  receives transaction  $T_1'$ ,  $N_2$  first determine whether  $T_1'$  is local transaction, if the local transaction, then go to step 3.2. When  $T_1'$  and  $T_2$  conflict, if the local transaction in the local implementation phase, undo  $T_2$  and identify  $T_2$  with 'abort' state. Go to step 3.3;

Step 3.2 Node  $N_1$  receives transaction  $T_1'$ ,  $T_1'$  is local transaction, redo  $T_1$  if  $T_1$  is on 'abort' state.

Step 3.3 Node sends successfully message (Executed) if transaction executed successfully, or sends the 'Abort' message.

**Step 4: Submission of global transaction**

If node  $N_1$  receives ‘Executed’ message from all members of the group, commits the transaction  $T_1'$ ; otherwise the message just received an ‘Abort’, rollback the transaction  $T_1'$ .

**5 Performance Testing**

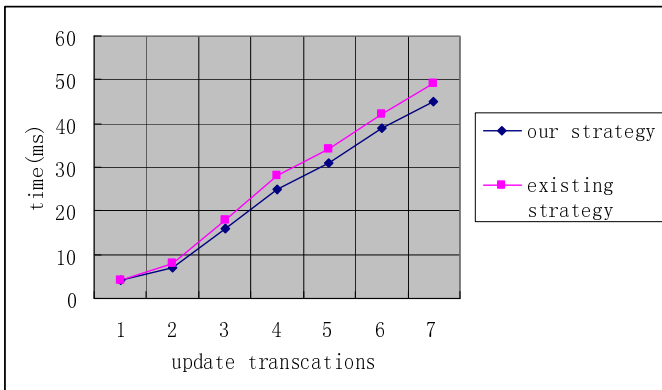
**5.1 Test Environment**

We implemented a simple communications support services based on group communication system Spread [7], and deployment of nodes in each group member. Simulation environment consists of three members of the PC as a group to communicate with each other, PC, the same configuration: CPU with Intel's Pentium4 processor, 1G RAM, 2.93GHz clock speed, 100G hard drive. Front-end using a PC to impersonate the client, the PC sends transactions to the member of group.

Test using the table name ‘table\_group’, a table field of type integer primary key is a 50-byte char type, an integer random number as a data type. Update transaction request to update the type of submission.

**5.2 Performance Test**

We simulate users submit continuous 30, 50, 100, 150, 200, 250, 300 update transactions request situations, observing the system time to complete transactions and copies of each node to the time required for data consistency.



**Fig. 5.** The time to complete transactions

Figure 5 shows that due to the atomic broadcast mechanism and removing the read-only transactions in global transactions, so the amount of data increases with the transaction, relative to the original replication strategy, the system significantly reduced traffic, the time to complete transactions also decreased significantly. Thus, a

large volume of transactions in the system case, this strategy used in the performance was significantly better than the existing replication strategies.

## 6 Conclusion

In this paper we describe data synchronization method of classification of distributed database system. The replication strategy for the current response time is too long, the shortcomings of the system communication consumption, we present a global transaction-oriented data synchronization strategy based on the atomic broadcast mechanism in group communication. The strategy guarantees the data consistency while reducing the consumption of system communication, and has been well verified in the experimental results. The next step will further improve the concurrent execution of transactions in the conflict resolution mechanisms in distributed database system, in order to achieve better data synchronization.

## References

1. Xiao, Y.-Y.: Technology of distributed real-time database. Science Press, Beijing (2009)
2. Gray, J., Helland, P., O'Neil, P., et al.: The Dangers of Replication and a Solution. In: ACM SIGMOD 1996, Montreal (1996)
3. Yao, L., Yang, H., Wang, Z., et al.: SyncML data synchronization protocol based on the ability to adapt to processing. *Computer Engineering* 35(5), 68–72 (2009)
4. Pedone, F., Schiper, A.: Using optimistic Atomic Broadcast in Transaction Processing Systems. Technical Report, Department of Computer Science (March 1999)
5. Kemme, B., Alonso, G.: Database Replication Based on Group Communication. Technical Report No. 289
6. Kemme, B., Alonso, G.: Don't be lazy, be consistent: Postgres-R, A new way to implement Database Replication. In: Int'l Conf. on Very Large Databases (2000)
7. Amir, Y., Nita-Rotaru, C., Stan Ton, J., et al.: Secure spread: an integrated architecture for secure group communication. *IEEE Trans. on Dependable and Secure Computing* (2005)



# Kinetic Study of Vanadium-Oxides' Magneto-sputter Deposition Based on Experimental Reaction Data

Zheng Cui, Yu Lu, Zuochun Shen, and Jianye Lu

National Key Laboratory of Science and Technology on Tunable laser,  
Harbin Institute of Technology, Harbin 150080, China

**Abstract.** Based on a physical model, the growth rates and the transformation rates in the magneto-sputter process for  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  are obtained by fitting the experimental data. From these values one can see, under the conditions of the experiment in this article, the growth rate for  $VO_2$  is higher than that for  $V_2O_3$  and  $V_2O_5$  and the transformation rate for  $VO_2$  transforming to  $V_2O_5$  is higher than that for  $V_2O_3$  transforming to  $VO_2$ . Also, in the view of the deposition time, in the beginning of deposition process, the percentage contents of each kind of vanadium-oxides change rapidly, and with the time growing, this change become gently. If both concerning the thickness of the deposited film and the percentage content of  $VO_2$ , under the conditions of the experiment in this paper, the deposition time should be round 20 minutes.

**Keywords:** physical model, vanadium-oxide, rate, fitting.

## 1 Introduction

The types of vanadium oxides are various. Besides the familiar  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$ , there also exist many other structures of vanadium oxides, such as  $V_6O_{13}$  etc. Usually, vanadium oxides can be expressed as  $V_xO_y$ . Among all these vanadium oxides,  $VO_2$  is the most attractive one. It is not only because of its good thermo-sensitivity, but also of its phase-transition point which approximates the room temperature most. In the single-crystal status,  $VO_2$  shows evident metal-semiconductor phase-transition property when the temperature rises to 341K. Along with the occurrence of phase-transition, many physical characteristics of  $VO_2$ , such as refractive index, reflectivity and resistance, change abruptly. The change rate of the resistance can be even as high as  $10^4$ - $10^5$  times. Electric-switches and opto-switches manufactured by utilizing these electric- and opto-characters can be used widely in the areas of temperature-sensing, changeable reflecting index mirror and laser protection [1-3] etc. At present, the studies on  $VO_2$  concentrate on how to obtain  $VO_2$  with high purity.

The purpose of the physical model prompted in this paper is to analysis the changes of percentage contents' for  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  during the process of depositing by magneto-sputter, in order to obtaining  $VO_2$  with high percentage content, when concerning both the thickness of the deposited film and the percentage content of  $VO_2$ . We hope these studies can be conducted to the improvement of  $VO_2$  thin film quality in the further annealing process.

## 2 Theoretical Descriptions

During the magneto-sputter process (also other thermal-vapor processes), the ways to generate vanadium-oxides' should be only two. One is to generate directly; another by transforming from other kinds of vanadium-oxides. For example,  $V_2O_5$  can be generated from  $VO_2$  by further reacting with oxygen etc. In order to simplify the following discussion, only  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  are considered in the system in the following. And also,

- I. Let  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  represent the possibilities of the direct generation rates of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  respectively;
- II. Let  $\eta_1$ ,  $\eta_2$  represent the transformation-rates of  $V_2O_3$  to  $VO_2$  and  $VO_2$  to  $V_2O_5$  separately;
- III. Let  $A(t)$ ,  $B(t)$ ,  $C(t)$  represent the total contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  at an arbitrary time  $t$  separately.

So, at any time point  $t$ , the changes of the contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  in the system can be expressed as:

$$dA(t) = \alpha_1 dt - \eta_1 A(t) dt \quad (1)$$

$$dB(t) = \alpha_2 dt + \eta_1 A(t) dt - \eta_2 B(t) dt \quad (2)$$

$$dC(t) = \alpha_3 dt + \eta_2 B(t) dt \quad (3)$$

Solve the three equations above, one can have:

$$A(t) = \frac{\alpha_1}{\eta_1} \{1 - \exp(-\eta_1 t)\} \quad (4)$$

$$B(t) = \frac{\alpha_1 + \alpha_2}{\eta_2} \cdot [1 - \exp(-\eta_2 t)] \quad (5)$$

$$- \frac{\alpha_1}{\eta_2 - \eta_1} \cdot [\exp(-\eta_1 t) - \exp(-\eta_2 t)]$$

$$C(t) = (\alpha_1 + \alpha_2 + \alpha_3) \cdot t + \frac{\alpha_1 + \alpha_2}{\eta_2} \cdot [\exp(-\eta_2 t) - 1] \quad (6)$$

$$- \frac{\alpha_1 \eta_2}{\eta_2 - \eta_1} \cdot \left[ \frac{\exp(-\eta_2 t) - 1}{\eta_2} - \frac{\exp(-\eta_1 t) - 1}{\eta_1} \right]$$

And, according to the equations (4-6), we can do the following discussions:

A. The relationship between the total contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  and the time  $t$ .

From (4-6), one can have:

$$\frac{dA}{dt} = \alpha_1 \exp(-\eta_1 t) \quad (7)$$

$$\frac{dB}{dt} = \frac{\alpha_1 \eta_1}{\eta_2 - \eta_1} \exp(-\eta_1 t) \quad (8)$$

$$- \frac{(\alpha_1 + \alpha_2) \cdot \eta_1 - \alpha_2 \eta_2}{\eta_2 - \eta_1} \exp(-\eta_2 t)$$

$$\frac{dC}{dt} = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{\alpha_1 \eta_2}{\eta_2 - \eta_1} \exp(-\eta_1 t) \quad (9)$$

$$+ \frac{(\alpha_1 + \alpha_2) \cdot \eta_1 - \alpha_2 \eta_2}{\eta_2 - \eta_1} \exp(-\eta_2 t)$$

I. It is clear from the equations (7-9) that the growth rate of the contents of  $V_2O_3$  and  $VO_2$  reduces rapidly to zero with the growth of the deposition time, in the mean time, the growth rate of  $V_2O_5$  becomes constant gradually;

II. The contents of  $V_2O_3$  and  $VO_2$  trend to be zero when the deposition time trends to infinity.

B. The relationship between the percentage content of  $VO_2$  and the deposition time  $t$ .

At each time point  $t$ , the percentage content of  $VO_2$  can be represented as :

$$P_B(t) = \frac{B(t)}{A(t) + B(t) + C(t)} = \frac{B(t)}{(\alpha_1 + \alpha_2 + \alpha_3) \cdot t} \quad (10)$$

So the following relation exists:

$$\frac{dP_B(t)}{dt} = \frac{F}{(\alpha_1 + \alpha_2 + \alpha_3) \cdot t^2} \quad (11)$$

Here,

$$\begin{aligned} F &= \beta_1 \eta_1 \exp(-\eta_1 t) \cdot t \\ &+ (\beta_1 \eta_1 - \alpha_2) \exp(-\eta_2 t) \cdot t - \beta_1 \exp(-\eta_1 t) \\ &+ (\beta_1 + \beta_2) \exp(-\eta_2 t) - \beta_2 \end{aligned} \quad (12)$$

$$\beta_1 = \frac{\alpha_1}{\eta_2 - \eta_1} \tag{13}$$

$$\beta_2 = \frac{\alpha_1 + \alpha_2}{\eta_2} \tag{14}$$

And one can have

I. When  $t$  is little enough, the exponential term in the equation (12) trends to 1, and equation (11) can be expressed as following approximately:

$$\frac{dP_B(t)}{dt} = \frac{2\beta_1\eta_1 - \alpha_2}{(\alpha_1 + \alpha_2 + \alpha_3) \cdot t} \propto t^{-1} \tag{15}$$

So, in the beginning of the deposition process, the percentage content of  $VO_2$  changes rapidly, almost in the rate of  $t^{-1}$ ;

II. When  $t$  is large enough, equation (11) becomes  $\frac{dP_B(t)}{dt} \rightarrow 0$ , i.e. the change of the percentage content of  $VO_2$  become flat gradually, in the rate of  $t^{-2}$ , and trend to zero.

### 3 Kinetic Analyses

Two pieces of vanadium-oxides films are deposited by magneto-sputter with time of 25min and 45min separately. The results of XPS test are shown in fig.1 and the percentage contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  obtained by fitting are  $(30.966 \pm 1.099)\%$ ,  $(49.368 \pm 1.211)\%$ ,  $(19.666 \pm 0.925)\%$ ;  $(18.985 \pm 1.318)\%$ ,  $(22.226 \pm 1.624)\%$ ,  $(58.788 \pm 1.867)\%$  separately. The reacting parameters fitted according to the formula above are:  $\alpha_1=2.55431$ ;  $\alpha_2=9.22236$ ;  $\alpha_3=0.04210$ ;  $\eta_1=0.00834$ ;  $\eta_2=0.05778$ .

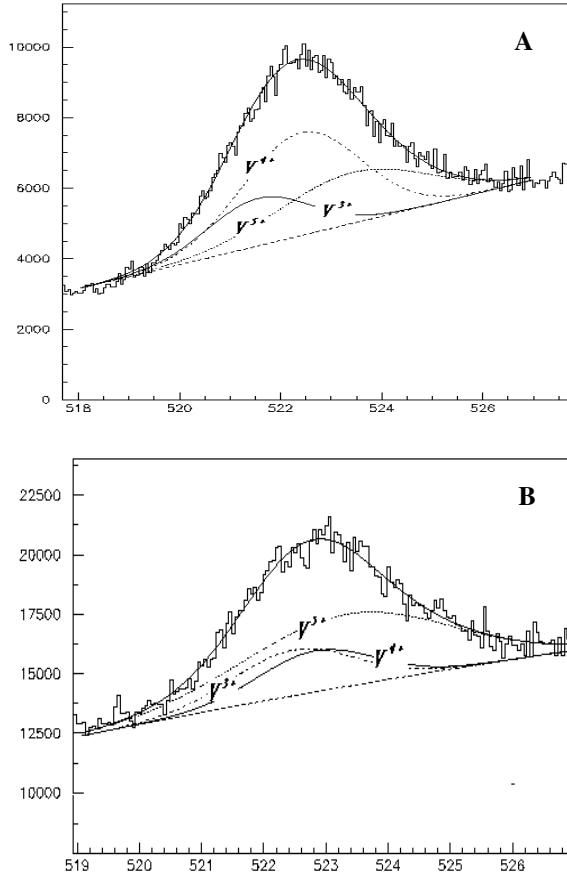
Furthermore, fig.2 shows the percentage contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  calculated according to equation (4-6). Then, one can have the following conclusions:

I. The direct generation rate of  $VO_2$  seems to be much greater than that of  $V_2O_3$ ,  $V_2O_5$ , considering  $\alpha_2 \gg \alpha_1 \gg \alpha_3$ ;

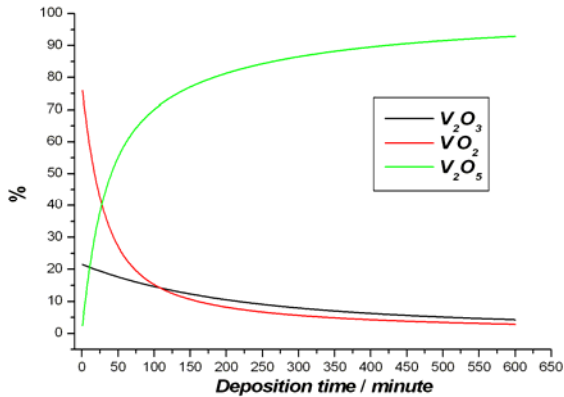
II.  $\eta_1 \ll \eta_2$  shows that the transformation rate from  $VO_2$  to  $V_2O_5$  is much higher than that from  $V_2O_3$  to  $VO_2$ ;

III. In the beginning of deposition process, the percentage content of  $VO_2$  is the highest (ref. Fig.2). But with increase of the deposition time, it reduces rapidly, meanwhile, the percentage content of  $V_2O_5$  increases rapidly. This period occurs at the first 50min of the deposition process, and with the continuous increase of time, the percentage contents of  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  become steady eventually;

IV. As we know, too thin film is too unsuitable for large crystal cell's generation to improve the  $VO_2$  film quality. But, long deposition time is harmful to increase the percentage content of  $VO_2$  in the deposited film. So, considering both two aspects above, it is important to determine the suitable deposition time. Under the conditions of the experiment in this paper, the best deposition time should be round 20min.



**Fig. 1.** XPS tests and fitted results for Vanadium-oxides deposited by magneto-sputter A: t=25min B: t=45min



**Fig. 2.** Fitted result: the change of the percentage contents of  $V^{3+}$ -,  $V^{4+}$ - and  $V^{5+}$ - oxide during the magnetron sputtered process

## 4 Conclusions

Based on a physical model, the growth rates and the transformation rates in magnetron sputter process for  $V_2O_3$ ,  $VO_2$  and  $V_2O_5$  are obtained by fitting the experimental data. From these parameters one can see that under the conditions of the experiment in this paper, the growth rate for  $VO_2$  is higher than that for  $V_2O_3$  and  $V_2O_5$  and the transformation rate for  $VO_2$  transforming to  $V_2O_5$  is higher than that for  $V_2O_3$  to  $VO_2$ . Also, in the view of deposition time, in the beginning of the deposition process, the percentage contents of each kind of vanadium oxides change rapidly, and become gently with the time growing. If concerning both the thickness of the deposited film and the percentage content of  $VO_2$ , under the conditions of the experiment in this article, the deposition time should be round 20 minutes.

**Acknowledgments.** Special thanks to the fund and many people who were involved in the National Key Laboratory of Science and Technology on Tunable laser, Harbin Institute of Technology Harbin for supporting this work.

## References

1. Hale, C.C.H., Orr, J.S., Gordon, H., et al.: Deposition and characterization of sputtered vanadium dioxide films. In: Proc. SPIE, vol. 1270, pp. 222–234 (1990)
2. Eden, D.D.: Some application involving the semiconductor-to-metal phase transition in  $VO_2$ . In: Proc. SPIE, vol. 185, pp. 97–102 (1979)
3. Baber, I., Trokman, S.: High-contrast optical storage in  $VO_2$  films. J. Appl. Phys. 46(5), 2111–2118 (1975)

# Cost Forecast of Overhead Transmission Line Based on BP Neural Network

Jun Zhou, Hanliang Wang, Yueguang Yang, and Jiangjun Ruan

College of Electrical Engineering, Wuhan University, Wuhan, Hubei Province, P.R. China  
Yyg0212@163.com.cn

**Abstract.** In order to make accurate and reasonable estimate of the overhead transmission line cost, which will become the basis for the establishment of project and cost, it is required to select an appropriate estimation method. Considering the traditional estimation methods' shortcomings of higher subjective nature and low forecast accuracy, this paper is based on the BP neural network, making use of the computer resources and proposes a new method of estimation using the specifications and performance parameters of transmission lines as training samples. By comparing the predicted value and actual value, it demonstrates that BP neural network can accurately predict the cost of overhead transmission lines so that the workload can be greatly reduced.

**Keywords:** BP neural network, overhead transmission lines, cost.

## 1 Introduction

In order to meet the demand of economic development in recent years, the power industry and electricity market grow fast as well. How to better control and reduce the project cost has become a problem for the owners of electricity constructions. The most effective measure is to predict the project cost quickly and accurately, which will become the basis for the establishment of project and cost control. However, the cost of erection of transmission lines is affected by many factors, and there are a lot of prediction parameters so it is difficult to establish a clear mapping function. Nowadays, the common estimation methods for cost are: fixed terms, empirical formula, fuzzy mathematics, etc., but subjective factors accounts for a large component in these types of methods, which results in inaccuracy in project estimation. Thus, this paper is based on the BP neural network, making use of the computer resources and proposes a new method of estimation using the specifications and performance parameters of transmission lines as training samples [1]. By comparing the predicted value and actual value, it demonstrates that BP neural network can accurately predict the cost of overhead transmission lines and the workload can be greatly reduced.

## 2 The Basic Principle of BP Neural Network

Neural network is a parallel, distributed processing architecture, which consists of processing units connected to the non-direction signal path [2]. It automatically forms

the decision-making region by its own learning mechanism, and it can achieve the highly nonlinear mapping between input and output data, by making best use of state information, which means separately dealing with the different state information, it can get the balanced convergence weight and mapping and can change its behavior according to the environment.

BP neural network is referred to widely used multi-level learning training algorithm for its broad applicability. It mainly achieves its prediction function by spreading the error response. Its structure consists of three neuron levels which are input layer, hidden layer and output layer, there are no links between neurons in layer, but neurons between layers connect with each other. BP neuron network model is shown in Fig 1.

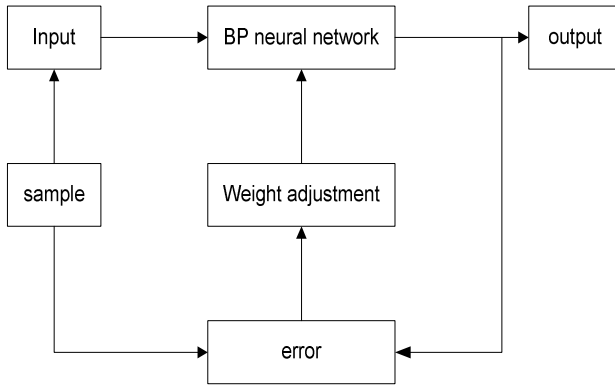


Fig. 1. BP neuron network model

The specific steps of algorithms can be summarized as follows:

- 1) Draw a sample from the training sample set, and put its input information into the network.

Let the number of the neurons  $x_i$  in the input layer is  $n$ , constituting  $n$  dimension input space.

- 2) Calculate each layer's node output

The number of the neurons in the hidden layer is  $z$ . When the threshold of the hidden unit  $j$  is  $\theta_j$ , the weight connecting to the input unit  $i$  is  $v_{ij}$ , and then the output of the hidden unit is:

$$h_j = f(\beta_j) = f\left(\sum_{i=0}^{n-1} v_{ij} x_i - \theta_j\right) \tag{1}$$

The number of the neurons in the output layer is  $m$ . When the threshold of the output unit  $k$  is  $\phi_k$ , the weight connecting to the input unit  $j$  is  $w_{jk}$ , and then the output of the hidden unit is:



$$y_k = f(\alpha_k) = f\left(\sum_{j=0}^{z-1} w_{kj}h_j - \phi_k\right) \tag{2}$$

3) Calculate the error between the actual output and predicted output

Let there is a kind of sample  $(X, T)$ , where  $T = [t_1, t_2, \dots, t_k, \dots]^T$  is the expected output of the input  $X$ . The actual output of the network is  $Y = [y_1, y_2, \dots, y_k, \dots]^T$ . We can make the network output  $Y = [y_1, y_2, \dots, y_k]^T$  approach to  $T = [t_1, t_2, \dots, t_k]^T$  by adjusting the weight.

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n) \tag{3}$$

$$v_{jj}(n+1) = v_{jj}(n) + \Delta v_{jj}(n) \tag{4}$$

4) Calculate from the output layer back to the first hidden layer; adjust the various network connection weights in the direction of reducing the error according to some certain rules.

We can define the value function reflecting the square sum of the output error:

$$E = \frac{1}{2} \sum_{k=0}^{m-1} (t_k - y_k)^2 \tag{5}$$

We can minimize  $E$  as soon as possible by gradient descent:

$$\Delta w_{jk}(n) = \eta \delta_k h_j \tag{6}$$

$$\Delta v_{jk}(n) = \eta \delta_k^* x_j \tag{7}$$

Where  $\eta$  is the learning rate. For the output layer unit, the error signal  $\delta_k$  is:

$$\delta_k = (t_k - y_k) y_k (1 - y_k) \tag{8}$$

If there is no special target for the hidden unit, the error signal  $\delta_k^*$  will be decided recursively by all those neurons  $x$  connecting directly to the neurons  $j$  and their weights.

$$\delta_k^* = h_j (1 - h_j) \sum_{k=0}^{m-1} \delta_k w_{jk} \tag{9}$$

5) Repeat the above steps on every sample in the training sample set until the entire training set error achieves the requirement [3-5].

### 3 The Transmission Line Cost Estimation Neural Network Design

1) The choice of network layer.

Without limiting the number of hidden nodes, the two layers BP network can implement any nonlinear mapping, taking into account that the sample of overhead transmission lines cost is relatively small; fewer hidden layer nodes can realize the hyperplane division sample space. Therefore, choosing the two layers BP network is appropriate.

**Table 1.** The parameters of the overhead transmission lines and the gray comprehensive correlation of the parameters

Type	LGJ-2*240	LGJ-2*300	LGJ-2*400	LGJ-2*500	Correlation
<b>Bundle conductors</b>	2	2	2	2	0.626
<b>Nominal cross-section Al(mm<sup>2</sup>)</b>	240	300	400	500	0.652
<b>Nominal cross-section steel(mm<sup>2</sup>)</b>	40	40	35	45	0.618
<b>Root number(Al)</b>	26	24	48	48	0.885
<b>Root number(steel)</b>	7	7	7	7	0.626
<b>Calculation cross-sectional area Al(mm<sup>2</sup>)</b>	238.85	300.09	390.88	488.58	0.656
<b>Calculation cross-sectional area steel (mm<sup>2</sup>)</b>	38.9	38.9	34.36	43.1	0.618
<b>Diameter (mm)</b>	21.66	23.94	26.82	30	0.851
<b>DC resistance (Ω)</b>	0.1209	0.09614	0.07389	0.05912	0.571
<b>Pull-off force (N)</b>	83370	92220	103900	128100	0.729
<b>Quality per unit (kg/km)</b>	964.3	1133	1349	1688	0.685
<b>Cost (ten thousand/km)</b>	55	61	68	75	

- 2) The selection of overhead transmission lines' cost factors. Based on the actual project and the degree of impact on the cost, we choose the parameters to build model.

In the BP neural network modeling, the choice of model parameters is crucial for the model. In this paper, gray correlation method is used to make the cost of the overhead transmission lines as the characteristic parameters of the sequence and make the factors as the sequence of related factors. And then we can calculate the correlation degree of the sequence of related factors and the characteristic sequence by the gray correlation method, where the resolution coefficient  $\zeta$  is 0.5.

According to the gray comprehensive correlation theory, we can choose the parameters which have higher correlation with the cost sequence of the overhead transmission lines. And then we can build model with these parameters. The parameters of the overhead transmission lines and the gray comprehensive correlation of the parameters are shown in Tab.1. From the table, we can choose the parameters of the nominal cross-section, aluminum root number, the calculation cross-sectional area (Al), diameter, the pull-off force and the length and quality per unit of the overhead transmission line by comparing.

- 3) The design of input and output layer.

According to the BP neural network design features, there are no contacts between neurons in each layer, so we use the overhead transmission lines cost factors as input layer. In this paper, the nominal cross-section, aluminum root number, the calculation cross-sectional area (Al), diameter, the pull-off force and quality per unit of the overhead transmission line are regarded as input layer, the input node number is 6. The output layer is the overhead transmission line cost, the node number is 1.

- 4) The number of hidden layer nodes.

Too many hidden layer nodes will cause the learning time too long; but lacking of hidden nodes will make poor fault tolerance, and low ability of identifying none-learning sample. Thus it is necessary to combine a number of factors to design. According to the experience, the confirmations of hidden layer nodes are designed based on the following formula:

$$n = \sqrt{n_i + n_o} + a \tag{10}$$

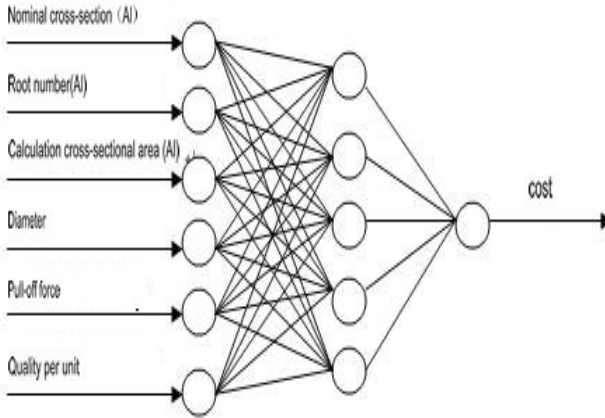
In this formula:  $n$  is the number of hidden nodes,  $n_i$  is the number of input nodes,  $n_o$  is the number of output nodes,  $a$  is a constant between 1 and 10.

- 5) Target selection.

The target output value is the actual cost of overhead transmission line.

- 6) The establishment of model.

In this BP neural network training process, you must first select a certain number of samples, and select the initial weights for learning, and then the output layer is the cost of overhead transmission line. Based on the above analysis, the overhead transmission line cost estimation model based on BP neuron network is shown in Fig 2.



**Fig. 2.** The overhead transmission line cost estimation model based on BP neuron network

7) Model calculation.

The calculation uses the Matlab software, the calculation process at first use the cost data of something which is similar to wire as the sample, trains the model, and adjusts the value, then use the model that has been trained to predict the cost of LGJ-2\*630 wire[6.7].

### 4 The Simulations of Overhead Transmission Lines' Cost

The specifications and performance parameters of different types of wire and its corresponding cost are shown in Tab. 2. If the 220 kV trend needs to send 300,000 kilowatts, it can choose the following wire cross-section of the wire.

**Table 2.** The specifications and performance parameters of different types of wire and its corresponding cost

Type	Nominal cross-section Al(mm <sup>2</sup> )	Root number (Al)	Calculati on cross-sectional area Al (mm <sup>2</sup> )	Diameter (mm)	Pull-off force (N)	Quality per unit (kg/km)	Cost (ten thous and/ km)
LGJ-2*240	240	26	238.85	21.66	83370	964.3	55
LGJ-2*300	300	24	300.09	23.94	92220	1133	61
LGJ-2*400	400	48	390.88	26.82	103900	1349	68
LGJ-2*500	500	48	488.58	30	128100	1688	75
LGJ-2*630	630	45	623.45	33.6	148700	2060	90

Using the five wire data as samples to train the model, and then use the trained model to predict cost of LGJ-2 \* 630 wire, the result is 84.7354, and the error is 5.85%. Therefore, BP neural network can conduct a relatively accurate cost prediction of overhead transmission lines, and can greatly reduce the workload.

## 5 Conclusion

### 1) Prospects

Through the simulation confirmations, it can be seen that BP neural network model based on types of wire to predict the cost of overhead transmission line is feasible. Though the data of overhead transmission line cost is not enough, which results in large forecast errors, the features of neural network such as strong ability of adaptation, none rejection to the new sample and the increasing predict accuracy as the sample grows, the BP neural network has a very broad application prospect in the cost prediction of overhead transmission line.

### 2) Suggestion

We should enhance the accumulation of data transmission lines, use BP neural network to estimate the purchase price of transmission lines on the basis of reliability, maintenance and the policy reserves of spare parts of the parameters. And then we can provide price proposals for the manufacturers of the transmission lines, give purchase price assessment reference to the power company which purchases transmission lines and provide technical support designed by cost to the research and design unit of transmission lines.



Zhou Jun was born in HuBei, China, on Nov 10, 1980. He received the B.S. degree from Wuhan University, Wuhan, China, in 2003, where he is currently working toward the Ph.D. degree of high-voltage and insulation technology in the School of Electrical Engineering. His major fields of interest are Power System Operation and Control and LCC technology.



Ruan Jiangjun was born in Zhejiang, China, on June 25, 1968. He received the B.S. and Ph.D. degrees in electric machine engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1990 and 1995, respectively. He finished his postdoctoral research from the Wuhan University of Hydraulic and Electrical Engineering, Wuhan, in 1998. He is currently a Professor with the School of Electrical Engineering, Wuhan University, Wuhan. His research interests include computational electromagnetics and high-voltage and insulation technology.

## References

1. Wen, X.: The Design to the Application of MATLAB Nerving Network. Beijing Science Press (2000) (in Chinese)
2. Zhang, W., Hua, X.: A method of Estimating the LCC of Ground Radar Equipment Based on BP Nerving Network. Journal of Air Force Engineering University (Natural Science Edition JV01) 10(1) (February 2009)
3. Guo, W.: Application of BP Neural Network Combining Forecast to Maintenance and Support Expense. Firepower and Command 29(3), 29–33 (2004) (in Chinese)
4. Zhu, H.: The Estimated Method and Application Based on the Dynamic BP Nerving Network. Computer Network 9(5), 11–16 (2006) (in Chinese)
5. Zhu, J.: Estimating Military Aircraft Cost Using Least Squares Support Vector Machines. International Journal of Plant Engineering and Management 6(2), 24–27 (2004)
6. Luo, Y.: The Method and the Application of the Equipment LCC. Ocean Press, Beijing (1992) (in Chinese)
7. Fu, C.: IFC Implementation in Lifecycle Costing. Journal of Harbin Institute of Technology 11(4), 437–441 (2004)

# Dynamic Aggregation Method of Induction Motors Based on Coherent Characteristics

Jun Zhou<sup>1</sup>, Jian Zhang<sup>2</sup>, Yueguang Yang<sup>1</sup>, and Yifeng Ju<sup>1</sup>

<sup>1</sup> College of Electrical Engineering, Wuhan University, Wuhan, Hubei Province, P.R. China

<sup>2</sup> Electric Power Research Institute of Jiangsu, Nanjing, Jiangsu Province, P.R. China  
Yyg0212@163.com.cn

**Abstract.** Parameters of induction motors have crucial effects on power system simulation. This paper proposes a dynamic aggregation method of induction motors. In this method, firstly, starting off the electromechanical models of induction motors, taking the coefficients of the electromechanical equations and steady slips of motors as the characteristic vectors, motors are grouped with C-means fuzzy clustering approach. Then, electro-mechanical equations of the equivalent motor is obtained by averaging that of the individual motors in the same group. Finally, parameters of equivalent motors are calculated based on the relationship of transient impedance before and after aggregation, and the principle that active and reactive power loads are equal before and after aggregation. Simulation results show that the proposed method can improve the precision of simulation.

**Keywords:** induction motor, coherent equivalence, C-means fuzzy clustering, dynamic aggregation.

## 1 Introduction

Model of each component is the base of power system simulation. Load model is the bottleneck of improving the precision of simulation due to its complexity. In load model, parameters of induction motor have crucial effects on voltage stability and angle stability.

In practice, thousands of induction motors scatter in a large distribution network. Simulation studies cannot be computationally feasible unless groups of individual induction motors be aggregated into a single equivalent motor.

There are mainly two approaches to build load model: the component-based approach and the measurement-based approach. The component-based method based on the investigation information, adopts aggregation algorithm to obtain the load model of a substation. There exist mainly five aggregation methods of induction motors: capacity weighted average method, power invariance method, steady state equivalent circuit of induction motor based method, dynamic equivalent circuit of induction motor based method, no load and block up working conditions of motor based method.

Reference [5] believes that load factor and critical slip have notable impact on transient stability of power system and an improved capacity weighted average aggregation method of induction motor is proposed. In reference [6,7], the steady state equivalent circuit of individual induction motor is transformed to the form of three paralleled impedances, based on which groups of motors are aggregated. This method can easily obtain parameters of the equivalent motor, but reactive load cannot be the same before and after aggregation. In reference [8], with the skin effect of motor taken into account, leakage inductances of stator and rotor are separated into two parts. Thus, impedance parameters of motor add to seven. In reference [9], an aggregation method of motors is proposed based on the no load and block up working conditions of motors. In reference [10], based on the simplified electromagnetic equations of motor an elegant static aggregation method of motors is proposed, but unfortunately, the stator resistance of motor is neglected.

All the methods mentioned above, belong to static aggregation method. The dynamic error is comparatively large. In this paper, based on their coherency, motors connected to the same bus are grouped by C-means fuzzy clustering method and a dynamic aggregation method of motors is proposed. The effectiveness of this method is validated by a simulation case.

## 2 Nomenclature

- $R_s$  Stator winding resistance of motor (p.u.).
- $X_s$  Stator leakage reactance of motor (p.u.).
- $X_m$  Magnetizing reactance of motor (p.u.).
- $R_r$  Rotor resistance of motor (p.u.).
- $X_r$  Rotor leakage reactance of motor (p.u.).
- $H$  Rotor inertia constant.
- $a$  Torque coefficient in proportion to square of speed.
- $b$  Torque coefficient in proportion to speed.
- $c$  Constant torque coefficient.
- $\omega$  Rotor speed (p.u.).
- $\omega_s$  Angle frequency of power system(rad/s).
- $I_d, I_q$   $d$ -axis and  $q$ -axis stator current.
- $V$  Bus voltage.
- $E'_d, E'_q$   $d$ -axis and  $q$ -axis transient EMF.
- $eq$  Subscript denoting parameters of equivalent motor.
- $i$  Subscript denoting parameters of  $i^{th}$  motor



### 3 Model of Induction Motor

#### Equivalent Circuit of Induction Motor

The steady state equivalent circuit of induction motor is shown as Fig.1, while the transient equivalent one is shown as Fig.2 [11].

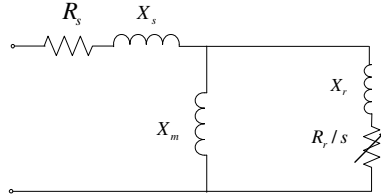


Fig. 1. Steady equivalent circuit of induction motor

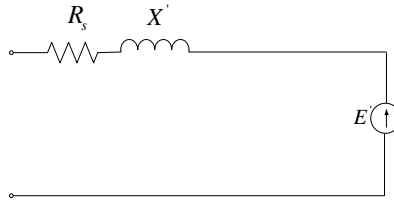


Fig. 2. Transient equivalent circuit of induction motor

According to [13], with the stator resistance neglected, the simplified electromechanical model of induction motor is:

$$T' dE'_q / dt = -E'_q + (X - X') / X V + T' \omega_s (\omega - 1) E'_d \tag{1}$$

$$T' dE'_d / dt = -E'_d - T' \omega_s (\omega - 1) E'_q \tag{2}$$

$$HX' d\omega / dt = VE'_d - T_m X' \tag{3}$$

Where,

$$T' = T'_0 X' / X$$

$$T'_0 = (X_r + X_m) / (\omega_s R_r)$$

$$X = X_s + X_m$$

$$X' = X_s + X_r X_m / (X_r + X_m)$$

$$T_m = a\omega^2 + b\omega + c = a(1-s)^2 + b(1-s) + c$$

#### Initialization of Induction Motor

According to Thevenin Theory, seen from rotor resistance to stator, the equivalent impedance of induction motor on steady state is:

$$\begin{aligned}
 Z_e &= R_e + jX_e = (R_s + jX_s) // jX_m + jX_r \\
 &= (R_s X_{sm}^2 - jX_p X_{rm}) / (X_p^2 + R_s^2 X_{sm}^2)
 \end{aligned}
 \tag{4}$$

Where

$$X_{sm} = X_s + X_m$$

$$X_{rm} = X_r + X_m$$

$$X_p = X_s X_m + X_r X_m + X_s X_r$$

$$R_e = R_s X_{sm}^2 / (X_p^2 + R_s^2 X_{sm}^2)$$

$$X_e = -X_p X_{rm} / (X_p^2 + R_s^2 X_{sm}^2)$$

The equivalent voltage is:

$$V_e = X_m / \sqrt{R_s^2 + X_{sm}^2} V \tag{5}$$

The electromagnetic torque of induction motor on steady state is:

$$T_e = \frac{V_e^2 (1-s) R_r / s}{\omega_s [(R_r / s + R_e)^2 + X_e^2]} \tag{6}$$

According to  $T_e = T_m$ , the steady state slip of rotor  $s_0$  can be obtained. The active and reactive power load of motor is:

$$P = \frac{V_e^2 R_r / s_0}{(R_r / s_0)^2 + X_e^2} \tag{7}$$

$$Q = \frac{V_e^2 X_e}{(R_r / s_0)^2 + X_e^2} \tag{8}$$

## 4 Motor Grouping Based on C-Means Fuzzy Clustering Method

### A. Introduction of C-Means Fuzzy Clustering Method

The idea behind C-means fuzzy clustering method is that it introduces the concept of membership from fuzzy mathematics. Against to ordinary C-means clustering method, C-means fuzzy clustering method can obtain the membership of each class for every sample. The advantage of C-means clustering method relies on that it reflects better the real world by expressing the intermediary property of sample.

The procedure of C-means fuzzy clustering method is as follows.

P1) According to the number of class pre-specified,  $c$  clustering centers are produced at random.

P2) Square of distance  $d_{ik}^2$  between each characteristic vector of sample  $x_k$  and each clustering center  $w_i$  is calculated as (9).

$$d_{ik}^2 = \|x_k - w_i\|_A^2 = (x_k - w_i)^T A(x_k - w_i) \tag{9}$$

Where, A is the weighting average matrix.

P3) suppose  $u_{ik}$  is the membership of  $i^{th}$  class for  $k^{th}$  sample, where  $u_{ik} \in [0,1]$  and  $\sum_{i=1}^c u_{ik} = 1$ . Real number  $m \in [0,2]$  is a weighting index. Objective function is formulated by (10).

$$\begin{aligned} \min J_m\{U, W\} &= \min \left\{ \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \right\} \\ &= \sum_{k=1}^n \min \left\{ \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \right\} \end{aligned} \tag{10}$$

P4) According to the clustering principle, following Lagrange function is formulated.

$$F = \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 + \lambda_k \left( \sum_{i=1}^c u_{ik} - 1 \right) \tag{11}$$

Where  $\lambda_k (k = 1, 2, \dots, n)$  is Lagrange multiplier. According to Kuhn-Tucker Theory, the necessary conditions of the objective function minimization are:

$$u_{ik} = 1 / \left[ \sum_{j=1}^c (d_{ik} / d_{jk})^{2/(m-1)} \right] \tag{12}$$

$$w_i = \sum_{k=1}^n (u_{ik})^m x_k / \sum_{k=1}^n (u_{ik})^m \tag{13}$$

By iteration, memberships and clustering centers are updated until difference between the latest two objective function values is lower than a threshold value or the most iteration number is reached.

### B. Choice of Characteristic Vector

Suppose there are k induction motors connected to the same bus. If the steady state slip  $s_0$ , steady state transient electric potential  $E_0'$  and the coefficients of the electromechanical equations of individual motor  $T'$ ,  $HX'$ ,  $(X - X')/X$ ,  $aX'$ ,  $bX'$ ,  $cX'$  are the same, then the k motors can be seen as only one motor. In practice, this condition cannot be satisfied, but motors of which those parameters are close can be classified into the same group and be seen as one motor. In this paper, for convenience, we choose  $(T', HX', (X - X')/X, aX', bX', cX', s_0)$  as the characteristic vector for motor grouping. Difference of per unit rotor speed  $\omega$  for motors in the same group is thought to be trivial and neglected, that is, motors in the same group are thought to be coherent.

## 5 Aggregation of Induction Motors

A. Aggregation of the Simplified Electromechanical Equations of Motors in the Same Group.

Averaging the electromechanical equations of motors in the same group we can obtain the electromechanical equations of the equivalent motor shown as (14)-(16).

$$\begin{aligned} d(\sum_{i=1}^n T_i' E_{qi}') / n / dt &= -(\sum_{i=1}^n E_{qi}') / n \\ +V[\sum_{i=1}^n (X_i - X_i') / X_i] / n &+ (\omega - \omega_s)(\sum_{i=1}^n T_i' E_{di}') / n \end{aligned} \quad (14)$$

$$\begin{aligned} d(\sum_{i=1}^n T_i' E_{di}') / n / dt &= \\ -(\sum_{i=1}^n E_{di}') / n &- (\omega - \omega_s)(\sum_{i=1}^n T_i' E_{qi}') / n \end{aligned} \quad (15)$$

$$\begin{aligned} (\sum_{i=1}^n H_i X_i') / n d\omega / dt &= \\ V(\sum_{i=1}^n E_{di}') / n &- (\sum_{i=1}^n T_{mi}' X_i') / n \end{aligned} \quad (16)$$

Equations (14)-(16) can be simplified to (17)-(19).

$$\begin{aligned} T_{eq}' dE_{q,eq}' / dt &= \\ -E_{q,eq}' + (X_{eq} - X_{eq}') / X_{eq} V &+ T_{eq}' (\omega - \omega_s) E_{d,eq}' \end{aligned} \quad (17)$$

$$T_{eq}' dE_{d,eq}' / dt = -E_{d,eq}' - T_{eq}' (\omega - \omega_s) E_{q,eq}' \quad (18)$$

$$\begin{aligned} H_{eq} X_{eq}' d\omega / dt &= \\ VE_{d,eq}' - (a_{eq} X_{eq}' \omega^2 &+ b_{eq} X_{eq}' \omega + c_{eq} X_{eq}') \end{aligned} \quad (19)$$

Where

$$E_{q,eq}' = (\sum_{i=1}^n E_{qi}') / n$$

$$E_{d,eq}' = (\sum_{i=1}^n E_{di}') / n$$

$$d(\sum_{i=1}^n T_i' E_{qi}') / n / dt \approx T_{eq}' dE_{q,eq}' / dt \quad (20)$$

$$d(\sum_{i=1}^n T_i' E_{di}') / n / dt \approx T_{eq}' dE_{d,eq}' / dt \quad (21)$$

$$T'_{eq} = (\sum_{i=1}^n T'_i) / n \tag{22}$$

$$(X_{eq} - X'_{eq}) / X_{eq} = [ \sum_{i=1}^n (X_i - X'_i) / X_i ] / n \tag{23}$$

$$H_{eq} X_{eq}' = (\sum_{i=1}^n H_i X'_i) / n \tag{24}$$

$$a_{eq} X_{eq}' = (\sum_{i=1}^n a_i X'_i) / n \tag{25}$$

$$b_{eq} X_{eq}' = (\sum_{i=1}^n b_i X'_i) / n \tag{26}$$

$$c_{eq} X_{eq}' = (\sum_{i=1}^n c_i X'_i) / n \tag{27}$$

Approximation effects caused by (20), (21) can be estimated by the following two equations.

$$\begin{aligned} & \left| \left( \sum_{i=1}^n T'_i E'_{qi} \right) / n - T'_{eq} E'_{q,eq} \right| / \left| T'_{eq} E'_{q,eq} \right| \leq \max \left| T'_i - (\sum_{i=1}^n T'_i) / n \right| \\ & \left| \left( \sum_{i=1}^n H_{eq} X'_{eq} \right) / n - H_{eq} X'_{q,eq} \right| / \left| H_{eq} X'_{q,eq} \right| \\ & \leq \max \left| H_i X'_i - (\sum_{i=1}^n H_i X'_i) / n \right| \end{aligned}$$

Thus, the variance of  $T'$ ,  $HX'$  in the same group smaller, the better of the approximation effect. The small variance of  $T'$ ,  $HX'$  in the same group can be guaranteed by the C-means fuzzy clustering method, according to optimization of the objective function.

### B. Calculation Method of Parameters for Equivalent Motor

According to Norton Theory, transient equivalent circuit of motors connected to the same bus is shown as Fig.3. The transient impedance of the equivalent motor can be obtained by (28).

$$R_{s,eq} + jX'_{eq} = 1 / \left( \sum_{i=1}^n 1 / (R_{st} + jX'_i) \right) \tag{28}$$

Substituting  $X'_{eq}$  from (28) into (22)-(27), the values of  $H_{eq}$ ,  $T'_{0,eq}$ ,  $X_{eq}$ ,  $a_{eq}$ ,  $b_{eq}$ ,  $c_{eq}$  can be obtained and (29)-(31) holds.

$$(X_{r,eq} + X_{m,eq}) / (\omega_s R_{r,eq}) = T'_{0,eq} \tag{29}$$

$$X_{s,eq} + X_{m,eq} = X_{eq} \tag{30}$$

$$X_{s,eq} + X_{r,eq} X_{m,eq} / (X_{r,eq} + X_{m,eq}) = X'_{eq} \tag{31}$$

Generally, magnetizing inductance  $X_{m,eq}$  is further larger than leakage inductance of stator and rotor  $X_{s,eq}$ ,  $X_{r,eq}$ . Thus,  $X_{s,eq}$ ,  $X_{r,eq}$ ,  $R_{r,eq}$  can be approximately calculated by (32)-(34).

$$X_{s,eq} = X_{eq} - \sqrt{X_{eq}(X_{eq} - X'_{eq})} \tag{32}$$

$$X_{r,eq} = X'_{eq} - X_{s,eq} \tag{33}$$

$$R_{r,eq} = X_{eq} / (\omega_s T'_{0,eq}) \tag{34}$$

Finally, shown as (35), (36), the magnetizing inductance  $X_{m,eq}$  and steady state slip  $s_{0,eq}$  of the equivalent motor can be obtained according to principle that the active and reactive power absorbed by motors before and after aggregated are equal.

$$P_{eq} = \sum_{i=1}^k P_i \tag{35}$$

$$Q_{eq} = \sum_{i=1}^k Q_i \tag{36}$$

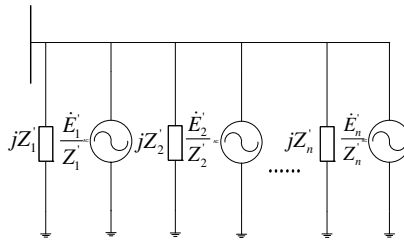


Fig. 3. Transient equivalent circuit of induction motors

## 6 Case Studies

### A. Grouping of Motors

Table 1 lists the characteristic vectors of the nine motors, parameters of which are indexed from [12]. The results of C-means fuzzy clustering are shown as Table 2. Obviously, the memberships of class one for motors NO.1 to NO.3 far outweighs that of class two. On the contrary, the memberships of class two for motors NO.4 to NO.9 far outweigh that of class one. So motors NO.1 to NO.3 belongs to class one, while motors NO.4 to NO.9 belongs to class two. Parameters of equivalent motors are shown as Table 3, where, M1 is the equivalent motor of the first three motors; M2 is the equivalent motor of the last six motors; M is the equivalent motor of the total nine motors.

**Table 1.** Characteristic vector of each motor

Motors	$T'$	$HX'$	$(X - X') / X$	$s_0$	$cX'$
1	0.00552728	0.103897	0.950787	0.03427	0.0611
2	0.00693847	0.166436	0.955437	0.02358	0.05380
3	0.00590828	0.22297	0.958979	0.01903	0.0375
4	0.0319691	0.22456	0.945262	0.003256	0.0342
5	0.0338508	0.177305	0.956791	0.001785	0.0202
6	0.040032	0.20259	0.961445	0.001113	0.0150
7	0.043005	0.240701	0.978083	0.001729	0.0254
8	0.0535508	0.223498	0.963103	0.001167	0.0211
9	0.0540806	0.192412	0.9662	0.001559	0.0284

**Table 2.** Membership of each motor

Motors	Class One	Class Two
1	0.9232	0.0768
2	0.9905	0.0095
3	0.6938	0.3062
4	0.0816	0.9184
5	0.0647	0.9353
6	0.0241	0.9759
7	0.0261	0.9739
8	0.0292	0.9708
9	0.0431	0.9569

**Table 3.** Parameters of equivalent motors

Motors	$M_1$	$M_2$	$M$
$R_s$	0.0397	0.0009818	0.00096767
$X_s$	0.1508	0.0062	0.0060
$X_m$	6.7153	0.3467	0.3295
$X_r$	0.1291	0.0007667	0.001
$R_r$	0.1473	0.0061	0.0058
$H$	0.2759	8.5472	8.2545
$s_0$	0.0226	0.0017	0.0024

**B. Simulation Results**

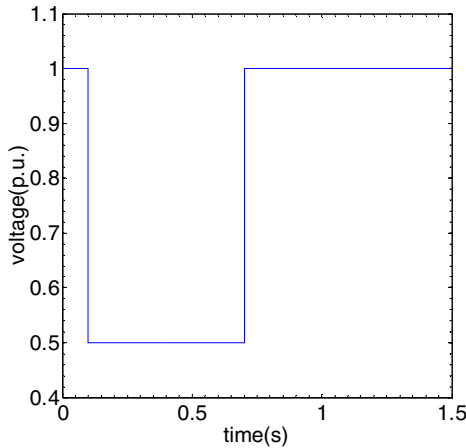
The first simulation is done in such a way that a voltage dip of 50% occurs at  $t = 0.1$  second, and is cleared later at  $t = 0.7$  second. Curves of active and reactive power absorbed by motors before and after aggregation are judging criteria of aggregation effect. Figures 4-8 show the dynamic response of three machines. The solid lines represent active or reactive power load before aggregation and the dashed lines represent active or reactive power load after aggregation.

Seen from Figs.5-6, it is observed that fitting effects with two equivalent motor far outweigh that with only one equivalent motor. This observation leads to the conclusion that it is not always advisable to represent all motors with a single equivalent one.

Curves of active and reactive power load of the first and second group of motors before and after aggregation are shown as Figs.7-8. It can be observed that the curves of active and reactive power before and after aggregation agree well.

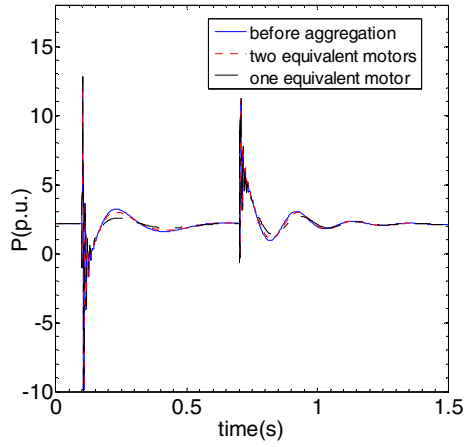
**C. Comparison with Selected Methods**

The second simulation is done in such a way that a voltage dip of 10% occurs at  $t = 0.1$  second, and is cleared later at  $t = 0.7$ . Curves of active and reactive power of motors before and after aggregation with proposed method, with capacity weighted average method [4] and with power invariance method [6, 7] are shown as Figs.9-10. From those figures, it is easy to observe that the fitting errors of active and reactive power with power invariance method are quite noticeable. The relative root mean square errors of active and reactive power by different methods are shown as Table 4. It is appropriate to reach the conclusion that the proposed method reduces the fitting errors of active and reactive power sharply.

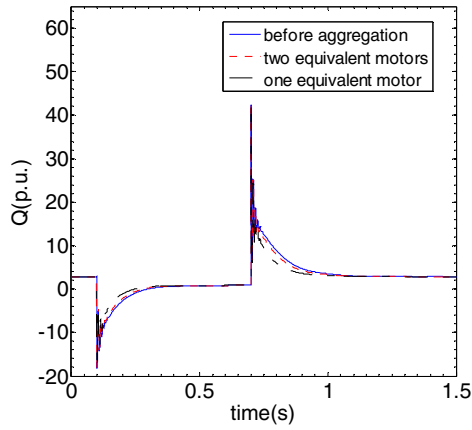


**Fig. 4.** Curve of interrupted voltage

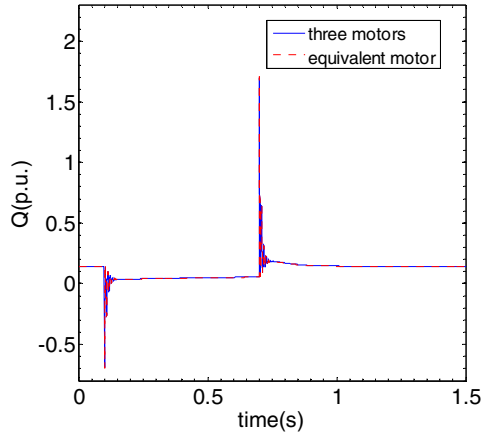




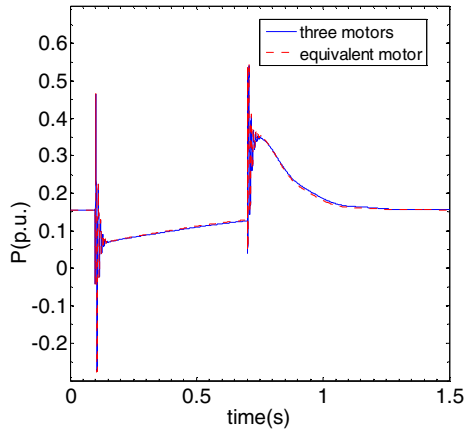
**Fig. 5.** Sum of active power of the 9 inductionmotors before and after aggregation



**Fig. 6.** Sum of reactive power of the 9 induction motors before and after aggregation



**Fig. 7.** Sum of active power of induction motors of the first group before and after aggregation



**Fig. 8.** Sum of reactive power of induction motors of the first group before and after aggregation

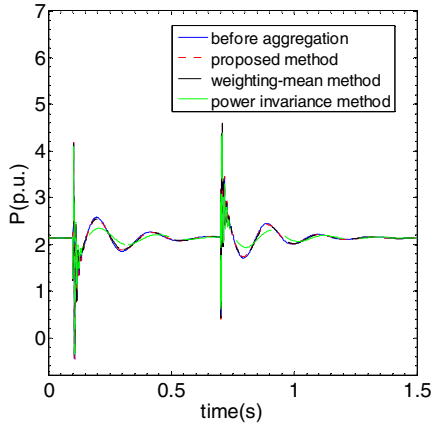


Fig. 9. Active power comparison with other methods

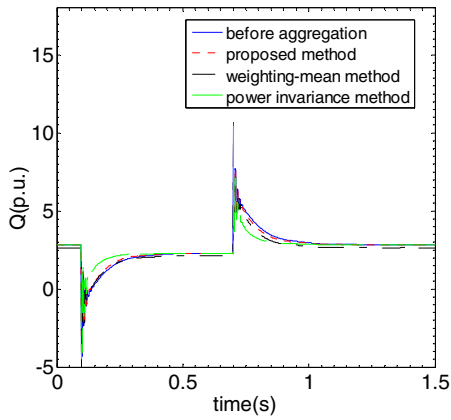


Fig. 10. Reactive power comparison with other methods

Table 4. Comparison of fitting errors with selected methods

Different Methods		Active Power Error	Reactive Power Error
Weighting Method	Average	0.0098	0.0743
		0.0404	0.1392
Power Invariance Method		0.0063	0.0282
Proposed Method			

## 7 Conclusion

This paper directly starts off the electro-mechanical equations of induction motor. With coefficients of electromechanical equations and steady state slip chosen as the characteristic vector, motors are grouped by C-means fuzzy clustering method. Motor in the same group are aggregated with dynamic aggregation method. Errors of active and reactive power reduce respectively from 0.0098, 0.0404 of capacity weighted average method and 0.0743, 0.1392 of power invariance method to 0.0063 and 0.0282.

### Acknowledgments

**Jun Zhou** was born in HuBei, China, on Nov 10, 1980. He received the B.S. degree from Wuhan University, Wuhan, China, in 2003, where he is currently working toward the Ph.D. degree of high-voltage and insulation technology in the School of Electrical Engineering. His major fields of interest are Power System Operation and Control and LCC technology.

**Jian Zhang** was born in Hubei Province, China, on October 19, 1983. He received the Ph.D. degree in Electric Engineering from Wuhan University, Wuhan, China. Now he is working at Electric Power Research Institute of Jiangsu in Nanjing. His research interests include electric load modeling.

### References

1. Cigre, D.: Task Force. Load modeling and dynamics. *Electra* 1(130), 124–141 (1990)
2. IEEE Task Force. Load representation for dynamic performance analysis. *IEEE Trans.on Power Systems* 8(2), 472–482 (1993)
3. IEEE Task Force Standard load models for power flow and dynamic performance simulation. *IEEE Trans. on Power Systems* 10(3), 1302–1313 (1995)
4. Ju, P., Ma, D.: Load modeling of power system, 2nd edn. China Electric Press, Beijing (2008)
5. Guo, J., Yu, Y., Ni, P., Zeng, Y.: An improved weighting mean aggregation method of induction motors considering load rates and critical slips. *Automation of Electric Power Systems* 32(14), 6–10 (2008)
6. Zhang, J., Zhang, C., Yan, A., Zhang, P., Li, K.: Aggregation of multi induction motors based on the self-organized neural network and steady state model. *Automation of Electric Power Systems* 31(11), 44–48 (2007)
7. Franklin, D.C., Amorelato: Improving dynamic aggregation of induction motor models. *IEEE Transactions on Power Systems* 9(4) (November 1994)
8. Akbaba, M., Fakhro, S.Q.: New model for single-unit representation of induction motor loads, including skin effect, for power system transient stability studies. In: *IEEE Proceedings-B*, vol. 139(6) (November 1992)
9. Zhang, H., Tang, Y., Zhang, D., et al.: Equivalent model for induction-motor load connected to high voltage bus. In: *Proceedings of the CSEE*, 2006, vol. 26(24), pp. 1–4 (2006) (in Chinese)

10. Taleb, M., Akbaba, M., Abdullah, E.A.: Aggregation of induction machines for power system dynamic studies. *IEEE Transactions on Power systems* 9(4) (November 1994)
11. Tang, Y., Hou, J., Liu, W.: The modeling of distribution network and var compensator and induction motor in the Load model for power system digital simulation. *Proceedings of the CSEE*, 2005 25(3), 8–12 (2005)
12. Cathey, J.J., Cavin, R.K., Ayoub, A.K.: Transient load model of an induction motor. *IEEE Transactions on Power Systems and Apparatus PAS-92*(4), 1399–1406 (1973)
13. Ahmed-Zaid, S., Taleb, M.: Structural modeling of small and large induction machines using integral manifolds. *IEEE Transactions on Energy Conversion* 6(3) (September 1991)

# Application of the Monopulse Tracking in Mimo Radar

Guodong Qin<sup>1,\*</sup> and Baixiao Chen<sup>2</sup>

<sup>1</sup> School of Electronic Engineering, Xidian Univ., Xi'an 710071, P.R. China

<sup>2</sup> National Key Lab of Radar Signal Processing, Xidian Univ. Xi'an 710071, P.R. China  
gdqin@mail.xidian.edu.cn

**Abstract.** MIMO Radar transmits the orthogonal signals, which results in the nondirectional of the pattern, the sum-difference beam can not be formed in radar receiver. In this paper, the azimuth and elevation sum-difference beam of the target corresponding to receiving station are formed using DBF(digital beam forming) technology, and range, the azimuth and elevation sum-difference beam of the target corresponding to transmitting station are formed using SIAR(synthetic impulse and aperture) technology. Therefore the error signal is formulated. And the tracking in MIMO radar is completed with a low computation complexity. The simulation results indicate the validity of the methods presented in this paper.

**Keywords:** MIMO radar, monopulse tracking, DBF, SIAR.

## 1 Introduction

Target tracking algorithms are mostly based on sub-space method in array signal processing, such as eigenvalue decompositions and singular value decompositions [1-4], stochastic gradient algorithms[5] and conjugate gradient method[6]. And all these algorithms suppose the angles of the targets vary with time rapidly. By contraries, the Maximum Likelihood Method[7] works when the angles vary with time slowly. All these algorithms have a high computation complexity, and can't be completed in a real-time system. The monopulse tracking[8] is applied in all kinds of systems because of the low computation complexity and high accuracy. MIMO Radar transmits the orthogonal signals, which results in the nondirectional of the pattern, the sum-difference beam can not be formed in radar receiver. In this paper, the range, azimuth and elevation sum-difference beam of the target are formed using DBF(digital beam forming) technology and SIAR[9](synthetic impulse and aperture) technology.

## 2 Signal Model

MIMO Radar uses a sparse-array of  $N_e$  elements to emit orthogonal waveforms, and a  $N_r$ -element uniform plane array to receive the echoes. Every element emits the FMCW (Frequency Modulated Continuous Wave) signals of different carrier frequencies.

---

\* Corresponding author: Guodong Qin, School of Electronic Engineering, Xidian Univ. He got the doctor degree in Xidian Univ. in 2009. His research interests including radar signal processing, array signal processing and target detection and tracking and so on.

The signal transmitted at the  $k$ th antenna in a pulse repetition interval can be expressed by

$$s_{ek}(t) = \text{rect}(t) e^{j2\pi(f_k t - 0.5\mu t^2)}, 0 \leq t < T_r, k = 1, \dots, N_e \tag{1}$$

Where function  $\text{rect}(t)$  denotes the gate signal,  $\text{rect}(t) = \begin{cases} 1, & 0 \leq t < T_e \\ 0, & T_e \leq t \leq T_r \end{cases}$ .  $f_k$  is the carrier frequency of the  $k$ th transmitting antenna unit,  $f_k = f_0 + \Delta f_k = f_0 + c_k \Delta f$ ,  $f_0$  is the center frequency, and  $\Delta f_k$  is the deviation of  $f_k$  from  $f_0$ ,  $c_k$  is the frequency codes,  $c_k \in \{-(N_e - 1)/2, \dots, (N_e - 1)/2\}$ ,  $\Delta f$  is frequency difference of transmitting antenna unit. Term  $\mu$  is the frequency sweep rate.  $B_u = \mu T_e$  denotes the signal bandwidth.  $T_e$  the pulse width,  $T_r$  the pulse repetition interval.

Without loss of generality, the signal decay factors corresponding to target RCS and propagation distance of each transmitted signal are assumed equal to each other, and they are omitted in the following deduction. Hence the received signals of  $l$ th receiving antenna is

$$\begin{aligned} s_{kl}(t) &= \sum_{k=1}^{N_e} s_{ek}(t - \tau_{kl}) \\ &= \sum_{k=1}^{N_e} \text{rect}(t - \tau_{kl}) e^{j2\pi[f_k(t - \tau_{kl}) - 0.5\mu(t - \tau_{kl})^2]}, k=1, \dots, N_e, l=1, \dots, N_r \end{aligned} \tag{2}$$

Where  $\tau_{kl}$  is the time delay for transmitted signal travelling from the  $k$ th transmitting unit to the target and then to the  $l$ th receiving antenna. While  $\tau_{kl} = \tau_0 - \Delta\tau_{tk} - \Delta\tau_{rl} - vt'/c$ ,  $c$  is the light speed.  $\tau_0 = 2R_0/c$ ,  $R_0$  is the initial range sum of the target corresponding to transmitting station and receiving station.  $\Delta\tau_{tk}$  is the time delay caused by wave path difference between the  $k$ th transmitting unit and the center of the transmitting array, and  $\Delta\tau_{rl}$  between the  $l$ th receiving unit and the center of the receiving array. In the case that the array aperture is not very large, i.e. the target range is much greater than the antenna aperture, for a array placed on the Y-Z plane,  $\Delta\tau_{tk} = (y_k \sin \theta_{T0} \cos \varphi_{T0} + z_k \sin \varphi_{T0})/c$   $\Delta\tau_{rl} = (y_l \sin \theta_{R0} \cos \varphi_{R0} + z_l \sin \varphi_{R0})/c$ , where  $\theta_{T0}$  and  $\varphi_{T0}$  is the azimuth and elevation of the target corresponding to transmitting station, respectively.  $\theta_{R0}$  and  $\varphi_{R0}$  are the azimuth and elevation of the target corresponding to receiving station, respectively.  $y_k, z_k$  and  $y_l, z_l$  are the coordinates of  $k$ th transmitting unit in transmitting array and  $l$ th receiving unit in receiving array, respectively. While  $v = 2v_r$ ,  $v_r$  is the radial velocity of the target. While  $t' = mT_r + t$  ( $0 < t \leq T_r$ ),  $t$  is the fast time,  $m$  is repetition period number index.

The received signal is demodulated by multiplying it with an ungated version of the transmit signal  $e^{j2\pi(f_0 t - 0.5\mu t^2)}$  and then sampled by an A/D converter. In order to separate the reflected signals with different carrier frequencies, with each carrier

frequency corresponding to different transmitting antenna, the sampled signal is passed through multiple channels, and multiplied by  $e^{-j2\pi\mu f_k t}$  in each channel separately, and then low-pass filtered individually. By neglecting the gate signal, the signal of the  $k$ th channel in  $l$ th receiving channel is given by

$$r_{kl}(t) = e^{j2\pi(\mu\tau_{kl}t - f_k\tau_{kl} - 0.5\mu\tau_{kl}^2)} \tag{3}$$

As shown in Equation (3), the range of the specific targets can be estimated from the frequency of demodulated and separated signals. This can be done by performing the fast Fourier transform (FFT) of a single pulse repetition interval, which is also known as range transform. Then the coherent integration signal of the  $k$ th channel in  $l$ th receiving channel is given by

$$r_{kl}(f_d, \tau) = e^{-j2\pi(f_k(\tau_0 - (\Delta\tau_k + \Delta\tau_r)))} e^{j\pi\mu T_e(\tau_k - \tau)} \cdot \frac{\sin \pi\mu T_e(\tau - \tau_k)}{\pi\mu(\tau - \tau_k)} \cdot e^{j\pi[(f_{dk} - f)(M-1)T_r]} \cdot \frac{\sin[\pi(f_{dk} - f_d)MT_r]}{\sin[\pi(f_{dk} - f_d)T_r]} \tag{4}$$

where  $f_{dk} = f_0 v/c + c_k \Delta f v/c$ ,  $\tau_k = \tau_0 + f_k v/c\mu$ ,  $M$  is the number of pulses used for coherent integration. And  $r_{kl}(f_{dk}, \tau_k)$  can be shown as

$$r_{kl}(f_{dk}, \tau_k) \approx e^{-j2\pi f_0 \Delta\tau_{rl}} e^{-j2\pi f_k(\tau_0 - \Delta\tau_k)} \tag{5}$$

Suppose  $R_0 = R_p + \Delta R_0$ ,  $\theta_{T0} = \theta_{Tp} + \Delta\theta_{T0}$ ,  $\varphi_{T0} = \varphi_{Tp} + \Delta\varphi_{T0}$ ,  $\theta_{R0} = \theta_{Rp} + \Delta\theta_{R0}$ ,  $\varphi_{R0} = \varphi_{Rp} + \Delta\varphi_{R0}$ .  $R_p$  is the rang sum of the target corresponding to transmitting station and receiving station.  $\theta_{Tp}$  and  $\varphi_{Tp}$  is the azimuth and elevation of the target corresponding to transmitting station, respectively.  $\theta_{Rp}$  and  $\varphi_{Rp}$  is the azimuth and elevation of the target corresponding to receiving station, respectively.  $\Delta R_0$  is range between the target and the target corresponding to the center of the range resolution cell.  $\Delta\theta_{T0}$  and  $\Delta\varphi_{T0}$  is the azimuth and elevation between the target and the target corresponding to the center of the azimuth and elevation resolution cell in the direction of transmitting station, respectively,  $\Delta\theta_{R0}$  and  $\Delta\varphi_{R0}$  denote that in the direction of receiving station. The purpose of the target tracking is finding  $(\Delta R_0, \Delta\theta_{T0}, \Delta\varphi_{T0}, \Delta\theta_{R0}, \Delta\varphi_{R0})$  using initial parameters  $(R_p, \theta_{Tp}, \varphi_{Tp}, \theta_{Rp}, \varphi_{Rp})$  in MIMO radar.

### 3 Target Tracking

The principle of the monopulse tracking is shown in Fig.1.  $\theta_0$  is the direction of the target,  $F(\theta + \delta)$ ,  $F(\theta - \delta)$  is the beam near the target. When  $\theta = \theta_0$ ,  $F(\theta + \delta) - F(\theta - \delta) = F(\theta_0 + \delta) - F(\theta_0 - \delta) = 0$ . If  $\theta \neq \theta_0$ , the error signal appears, and then the direction of the target  $\theta_0$  can be obtained.



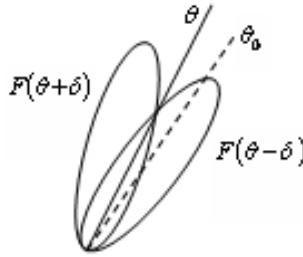


Fig. 1. The principle of the monopulse tracking

In MIMO radar, DBF and SIAR can be completed with the parameters  $(R + \delta_R, \theta_R + \delta_{\theta_R}, \varphi_R + \delta_{\varphi_R}, \theta_T + \delta_{\theta_T}, \varphi_T + \delta_{\varphi_T})$ ,  $(R - \delta_R, \theta_R - \delta_{\theta_R}, \varphi_R - \delta_{\varphi_R}, \theta_T - \delta_{\theta_T}, \varphi_T - \delta_{\varphi_T})$ . And then the beam  $F(R + \delta_R, \theta_R + \delta_{\theta_R}, \varphi_R + \delta_{\varphi_R}, \theta_T + \delta_{\theta_T}, \varphi_T + \delta_{\varphi_T})$  and  $F(R - \delta_R, \theta_R - \delta_{\theta_R}, \varphi_R - \delta_{\varphi_R}, \theta_T - \delta_{\theta_T}, \varphi_T - \delta_{\varphi_T})$  are obtained. According to monopulse tracking algorithms, the sum-difference beam are formed. And the target tracking in MIMO radar is shown in Fig.2.

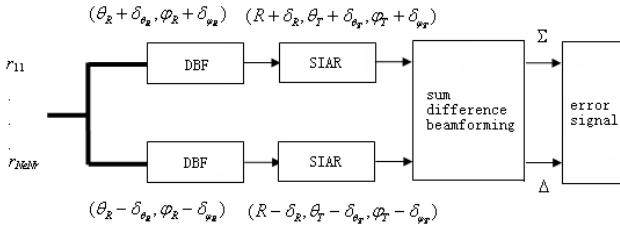


Fig. 2. Target tracking in MIMO radar

Suppose  $R = R_p + \Delta R$ ,  $\theta_R = \theta_{Rp} + \Delta\theta_R$ ,  $\varphi_R = \varphi_{Rp} + \Delta\varphi_R$ ,  $\theta_T = \theta_{Tp} + \Delta\theta_T$ ,  $\varphi_T = \varphi_{Tp} + \Delta\varphi_T$ . Then

$$\begin{aligned}
 & F(R \pm \delta_R, \theta_R \pm \delta_{\theta_R}, \varphi_R \pm \delta_{\varphi_R}, \theta_T \pm \delta_{\theta_T}, \varphi_T \pm \delta_{\varphi_T}) \\
 &= \exp(j\psi_0) \sum_{l=1}^{N_r} \exp(j2\pi f_0 (\tau_{rl} - \tau'_{rl})) \cdot \\
 & \sum_{k=1}^{N_r} \exp(j2\pi f_k (\tau_{ek} - \tau'_{ek})) \exp(j\xi_k (\Delta R - \Delta R_0)) \cdot \\
 & \exp(\mp j\xi_k \delta_R)
 \end{aligned} \tag{6}$$

where  $\psi_0 = -2\pi f_0 \tau_0$ ,  $\xi_k = 2\pi c_k \Delta f / c$ .

$$\Delta \tau'_{ik} = y_k \sin(\theta_T \pm \delta_{\theta_T}) \cos(\varphi_T \pm \delta_{\varphi_T}) / c + z_k \sin(\varphi_T \pm \delta_{\varphi_T}) / c$$

$\Delta \tau'_{rl} = y_l \sin(\theta_R \pm \delta_{\theta_R}) \cos(\varphi_R \pm \delta_{\varphi_R}) / c + z_l \sin(\varphi_R \pm \delta_{\varphi_R}) / c$ .  $\delta_R$  is half of the range cell.  $\delta_{\theta_k}$  and  $\delta_{\varphi_k}$  is half-power beamwidth corresponding to receiving station,

respectively.  $\delta_{\theta_r}$  and  $\delta_{\varphi_r}$  is half-power beamwidth corresponding to transmitting station, respectively.

**Range Tracking**

Suppose the angles of the target are known, Equation (6) can be given as

$$F(\Delta R \pm \delta_R) = N_r \exp(j\psi_0) \sum_{k=1}^{N_e} \exp(j\xi_k (\Delta R - \Delta R_0)) \cdot \exp(\mp j\xi_k \delta_R) \tag{7}$$

The sum beam is

$$\begin{aligned} \sum_R(\Delta R) &= F(\Delta R - \delta_R) + F(\Delta R + \delta_R) \\ &= 2N_r \exp(j\psi_0) \sum_{k=1}^{N_e} \cos(\xi_k \delta_R) \exp(j\xi_k (\Delta R - \Delta R_0)) \end{aligned} \tag{8}$$

The difference beam is

$$\begin{aligned} \Delta_R(\Delta R) &= F(\Delta R + \delta_R) - F(\Delta R - \delta_R) \\ &= j2N_r \exp(j\psi_0) \sum_{k=1}^{N_e} \sin(\xi_k \delta_R) \exp(j\xi_k (\Delta R - \Delta R_0)) \end{aligned} \tag{9}$$

The error signal is given as[9]

$$E_R(\Delta R) = (\sum_R \cdot \Delta_R^* + \sum_R^* \cdot \Delta_R) / |\sum_R|^2 \tag{10}$$

Where \* denotes conjugate,  $|\cdot|$  denotes module. Apply Equation(8) and (9) in Equation(10), the error signal is given as

$$as \ E_R(\Delta R) = K_R \cdot (\Delta R - \Delta R_0) \tag{11}$$

Where  $K_R = \frac{2\delta_R}{N_e} \sum_{k=1}^{N_e} \xi_k^2$ , which is slope of the error signal. When  $\Delta R = \Delta R_0$ ,  $E_R(\Delta R) = 0$ , and  $\Delta R_0$  can be obtained.

**Angles Tracking**

Firstly, let us discuss the angles tracking corresponding to receiving station. Suppose  $\Delta R_0 = \Delta R$ ,  $\Delta\theta_{T0} = \Delta\theta_T$ ,  $\Delta\varphi_{T0} = \Delta\varphi_T$ , and DBF is done on the direction of  $(\theta_R + \delta_{\theta_r}, \varphi_{Rp})$  and  $(\theta_R - \delta_{\theta_r}, \varphi_{Rp})$ , the azimuth beam is given as

$$F(\theta_R \pm \delta_{\theta_r}, \varphi_{Rp}) = N_e \exp(j\psi_0) \sum_{l=1}^{N_r} \exp(j2\pi f_0 (\tau_{rl} - \tau'_{rl-\theta})) \tag{12}$$

Where  $\tau'_{rl-\theta} = (y_l \sin(\theta_R \pm \delta_{\theta_r}) \cos \varphi_{Rp} + z_l \sin \varphi_{Rp}) / c$ .

The sum-difference beam is given as

$$\begin{aligned} \Sigma_a(\theta_R) &= F(\theta_R - \delta_{\theta_R}) + F(\theta_R + \delta_{\theta_R}) \\ &\approx 2N_e \exp(j\psi_0) \sum_{l=1}^{N_r} \cos(\delta_{\theta_R} \alpha_{\theta R}) (1 + j(\alpha_{\theta R1}(\Delta\theta_{R0} - \Delta\theta_R) + \beta_{\theta R} \Delta\varphi_{R0})) \end{aligned} \tag{13}$$

$$\begin{aligned} \Delta_a(\theta_R) &= F(\theta_R + \delta_{\theta_R}) - F(\theta_R - \delta_{\theta_R}) \\ &\approx -2jN_e \exp(j\psi_0) \cdot \\ &\quad \sum_{l=1}^{N_r} \sin(\delta_{\theta_R} \alpha_{\theta R}) (1 + j(\alpha_{\theta R1}(\Delta\theta_{R0} - \Delta\theta_R) + \beta_{\theta R} \Delta\varphi_{R0})) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \alpha_{\theta R} &= 2\pi f_0 y_l \cos(\varphi_{Rp}) (\cos(\theta_{Rp}) - \sin(\theta_{Rp}) \Delta\theta_R) / c \\ \alpha_{\theta R1} &= 2\pi f_0 y_l \cos(\varphi_{Rp}) \cos(\theta_{Rp}) / c \\ \beta_{\theta R} &= \frac{2\pi f_0 (z_l \cos(\varphi_{Rp}) - y_l \sin(\varphi_{Rp}) (\sin(\theta_{Rp}) + c \cos(\theta_{Rp}) \Delta\theta_{R0}))}{c} \end{aligned}$$

The error signal is shown approximately

$$E_a(\Delta\theta_R) \approx K_{\theta_R} \cdot (\Delta\theta_R - \Delta\theta_{R0}) + \Theta_{\theta_R} \tag{15}$$

where  $K_{\theta_R} = \frac{2\delta_{\theta_R}}{N_r^2} \left( \sum_l \alpha_{\theta R} \alpha_{\theta R1} - \sum_l \sum_{l'} \alpha_{\theta R} \alpha'_{\theta R1} \right)$ , which is slope of the error signal.

$$\alpha'_{\theta R1} = 2\pi f_0 y_{l'} \cos(\varphi_{Rp}) \cos(\theta_{Rp}) / c, \quad \Theta_{\theta_R} = \frac{2\delta_{\theta_R}}{N_r^2} \left( \sum_l \alpha_{\theta R} \beta_{\theta R} - \sum_l \sum_{l'} \alpha_{\theta R} \beta'_{\theta R} \right) \Delta\varphi_{R0},$$

where  $\beta'_{\theta R} = \frac{2\pi f_0}{c} (z_{l'} \cos(\varphi_{Rp}) - y_{l'} \sin(\varphi_{Rp}) \sin(\theta_{Rp}) + \cos(\theta_{Rp}) \Delta\theta_{R0})$ ,  $l' = 1, \dots, N_r$ .

$\Theta_{\theta_R}$  is intercept, which is so small that can be ignored.

Similarly, DBF is done on the direction of  $(\theta_{Rp}, \varphi_R + \delta_{\varphi_R})$  and  $(\theta_{Rp}, \varphi_R - \delta_{\varphi_R})$ , the elevation error signal is shown

$$E_e(\Delta\varphi_R) \approx K_{\varphi_R} \cdot (\Delta\varphi_R - \Delta\varphi_{R0}) + \Theta_{\varphi_R} \tag{16}$$

where  $K_{\varphi_R}$ , which is slope of the error signal.  $\Theta_{\varphi_R}$  is intercept, which is so small that can be ignored.

$$\begin{aligned} K_{\varphi_R} &= \frac{2\delta_{\varphi_R}}{N_r^2} \left( \sum_l \alpha_{\varphi R} \alpha_{\varphi R1} - \sum_l \sum_{l'} \alpha_{\varphi R} \alpha'_{\varphi R1} \right) \\ \Theta_{\varphi_R} &= \frac{2\delta_{\varphi_R}}{N_r^2} \left( \sum_l \alpha_{\varphi R} \beta_{\varphi R} - \sum_l \sum_{l'} \alpha_{\varphi R} \beta'_{\varphi R} \right) \Delta\theta_{R0} \end{aligned}$$

where

$$\alpha'_{\varphi_{R1}} = \frac{2\pi f_0 [z_l \cdot \cos(\varphi_{Rp}) - y_l \cdot \sin(\theta_{Rp}) \cos(\varphi_{Rp})]}{c}, l' = 1, \dots, N_r$$

$$\beta'_{\varphi_R} = \frac{2\pi f_0 (y_l \cdot \cos(\theta_{Rp})(\cos(\varphi_{Rp}) - \sin(\varphi_{Rp})\Delta\varphi_{R0}))}{c}, l' = 1, \dots, N_r.$$

Secondly, let us discuss the angles tracking corresponding to transmitting station. Suppose  $\Delta R_0 = \Delta R$ ,  $\Delta\theta_{R0} = \Delta\theta_R$ ,  $\Delta\varphi_{R0} = \Delta\varphi_R$ . SIAR is done on the direction of  $(\theta_T + \delta_{T_r}, \varphi_{T_p})$  and  $(\theta_T - \delta_{\theta_r}, \varphi_{T_p})$ , the azimuth beam is given as

$$F(\theta_T \pm \delta_{\theta_r}, \varphi_{T_p}) = N_r \exp(j\psi_0) \sum_{k=1}^{N_e} \exp(j2\pi f_k (\tau_{ek} - \tau'_{ek-\theta})) \quad (17)$$

Where  $\tau'_{ek-\theta} = (y_k \sin(\theta_T \pm \delta_{\theta_r}) \cos \varphi_{T_p} + z_k \sin \varphi_{T_p}) / c$ .

The error signal is shown

$$E_a(\Delta\theta_T) \approx K_{\theta_r} \cdot (\Delta\theta_T - \Delta\theta_{T0}) + \Theta_{\theta_r} \quad (18)$$

where  $K_{\theta_r} = \frac{2\delta_{\theta_r}}{N_e^2} \left( \sum_k \alpha_{\theta T} \alpha'_{\theta T1} - \sum_k \sum_{k'} \alpha_{\theta T} \alpha'_{\theta T1} \right)$

$$\alpha'_{\theta T1} = 2\pi f_k y_k \cdot \cos(\varphi_{T_p}) \cos(\theta_{T_p}) / c \quad \Theta_{\theta_r} = \frac{2\delta_{\theta_r}}{N_e^2} \left( \sum_k \alpha_{\theta T} \beta_{\theta T} - \sum_k \sum_{k'} \alpha_{\theta T} \beta'_{\theta T} \right) \Delta\varphi_{T0}.$$

Similarly, the elevation error signal is shown

$$E_e(\Delta\varphi_T) \approx K_{\varphi_r} \cdot (\Delta\varphi_T - \Delta\varphi_{T0}) + \Theta_{\varphi_r} \quad (19)$$

where

$$K_{\varphi_r} = \frac{2\delta_{\varphi_r}}{N_e^2} \left( \sum_k \alpha_{\varphi T} \alpha'_{\varphi T1} - \sum_k \sum_{k'} \alpha_{\varphi T} \alpha'_{\varphi T1} \right)$$

$$\Theta_{\varphi_r} = \frac{2\delta_{\varphi_r}}{N_e^2} \left( \sum_k \alpha_{\varphi T} \beta_{\varphi T} - \sum_k \sum_{k'} \alpha_{\varphi T} \beta'_{\varphi T} \right) \Delta\theta_{T0}$$

where

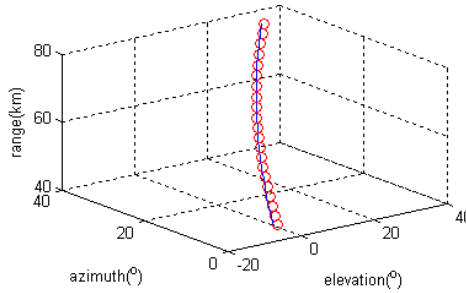
$$\alpha'_{\varphi T1} = 2\pi f_k [z_k \cdot \cos(\varphi_{T_p}) - y_k \cdot \sin(\theta_{T_p}) \cos(\varphi_{T_p})] / c, k' = 1, \dots, N_e$$

$$\beta'_{\varphi T} = \frac{2\pi f_k (y_k \cdot \cos(\theta_{T_p})(\cos(\varphi_{T_p}) - \sin(\varphi_{T_p})\Delta\varphi_{T0}))}{c}, k' = 1, \dots, N_e.$$

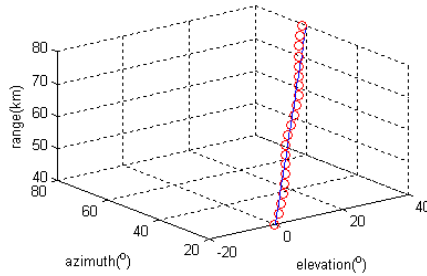
## 4 Simulation Results

The following system parameters are used in simulations.  $f_0 = 3\text{GHz}$ ,  $T_r = 512\mu\text{s}$ ,  $T_e = 450\mu\text{s}$ ,  $\Delta f = 2\text{MHz}$ ,  $B_u = 2\text{MHz}$ ,  $N_e = 32$ ,  $N_r = 16$ ,  $M = 128$ . The initial rang of the target is 40km, the azimuth and elevation corresponding to receiving station is  $20^\circ$  and  $0^\circ$ , respectively. The length of base line is 20km.

According to the geometric relationships of the radar, the azimuth and elevation corresponding to transmitting station is  $20^\circ$  and  $0^\circ$ , respectively. The target moves from the initial position at a constant speed, and the tracking trajectory is shown in Fig.3. The line is the real trajectory of the target, and the circle is tracking trajectory. It is shown that the method in this paper can track target effectively.



(a) Tracking trajectory of the target corresponding to transmitting station



(b) Tracking trajectory of the target corresponding to receiving station

**Fig. 3.** Tracking trajectory of the target

## 5 Conclusion

In this paper, monopulse tracking is applied in target tracking in MIMO radar. The range, azimuth and elevation sum-difference beam is formed using DBF and SIAR technology. And the error signal of range, azimuth and elevation is obtained. The target tracking is completed in MIMO radar with a low computation complexity and high accuracy.

**Acknowledgments.** This work is supported in parts by Professor Chen in National key lab of radar signal processing.

## References

1. Duan, M.-D., Gao, Z.-B., Ma, W., Li, J.-S.: Application of RE in the Development of Tractor Covering Parts. *Tractor & Farm Transporter*, 86–88 (2007)
2. Guo, L.-G., Chen, Y., Liu, X.-W.: Study of the Application of the Reverse Engineering Technology in the Design of the Complex Cavity-Surface Dies & Moulds. *Die & Mould Industry*, 12–17 (January 2002)
3. Mooen, M., Van Dooren, P., Vandwalle, J.: Parallel and adaptive high resolution direction finding. In: *Proc. SPIE Advanced Algorithm and Architectures for Signal Processing*, pp. 219–230 (1992)
4. Mooen, M., Van Dooren, P., Vandwalle, J.: Updating singular value decompositions: a parallel implementation. In: *SPIE Advanced Algorithm and Architectures for Signal Processing*, pp. 80–91 (1992)
5. Mooen, M., Van Dooren, P., Vandwalle, J.: A singular value decomposition updating algorithm for subspace tracking. *SIAM J. Matrix and Appl.* 13(4), 1015–1038 (1992)
6. Mooen, M., Van Dooren, P., Vandwalle, J.: A systolic array for SVD updating. *SIAM J. Matrix and Appl.* 14(2), 353–371 (1993)
7. Thompson, P.A.: An adaptive spectral analysis technique for unbiased frequency estimation in the presence of white noise. In: *Proc. 13th Asilomar Conf. Circuits. Syst. Comput. C*, Pacific Grove, pp. 529–533 (1980)
8. Fu, W., Dowling, E.M.: Conjugate gradient eigenstructure tracking for adaptive spectral estimation. *IEEE Transaction on Signal Processing* 43(5), 1151–1160 (1995)
9. Rao, C.R.: Tracking the direction of arrival of multiple moving targets. *IEEE Transaction on Signal Processing* 42(5), 1133–1144 (1994)
10. Mahafza, B.R., Huntsville, Alabama: Radar system analysis and design using MATLAB, pp. 385–392. HAPMAN&HALL/CRC (2000)
11. Baixiao, C., Shouhong, Z., Yajun, W., Jun, W.: Analysis and experimental results on sparse array synthetic impulse and aperture radar. In: *Proc. 2001 CIE Int. Conf. Radar*, pp. 76–80 (2001)

# A Class of Differential 4 Uniform Functions from Gold Functions

Dabin Zheng<sup>1,2</sup>

<sup>1</sup> LOIS, Graduate School of Chinese Academy of Sciences, Beijing 100049, China

<sup>2</sup> Faculty of Mathematics and Computer Science, Hubei University, Wuhan 430062, China  
zhengdabin@mmrc.iss.ac.cn

**Abstract.** It is an interesting and challenging topic in cryptography to construct Boolean (or vector Boolean) functions with various of good cryptographic properties. In this paper, by analyzing of a class of linearized equations, we prove that the differential uniformity of the functions constructed by exchanging two values of the Gold functions is 4, and the derived functions also have high algebraic degree and high nonlinearity.

**Keywords:** Boolean function, vector Boolean function, differential uniformity, algebraic degree, nonlinearity.

## 1 Introduction

Cryptographic functions usually are a core component of symmetric cryptographic algorithms [9]. In order to resistance to various known attacks, cryptographic functions used in cryptosystem must have a variety of good cryptographic properties. For example, the functions used in Block ciphers must have high nonlinearity to resist linear attacks [15], and low differential uniformity to resist differential attacks[3,16], and high algebraic degree to resist high order differential attacks[14]. For fields of even characteristic the lowest possible differential uniformity is 2, and the functions with this property are called APN (almost perfect nonlinear) functions. It is known that there are six classes of known APN power functions [2,10,11,13,17] and many APN polynomials [4,6,7,8].

It is almost impossible to find the cryptographic function which have all optimal cryptographic properties since they are usually mutual restraint. For instance, the AES (advanced encryption standard) uses a differentially 4 uniform function, but not a differentially 2 uniform function. So, it is meaningful to find more differential 4 uniform functions with other good cryptographic properties. Recently, there are some works on the differential 4 uniform functions [5,18,19]. The Gold APN function has the optimal differential uniformity but its algebraic degree is also low. In this paper, we construct a class of differential 4 uniform functions with high algebraic degree and high nonlinearity by exchanging two points in the value table of the Gold APN functions. According to analysis of the number of solutions of a class of linearized equations, we prove that the differential uniformity of the derived functions by exchanging two values of the Gold functions must be 4.

The rest of the paper is organized as follows. In section 2, some definitions and preliminaries are introduced. In section 3, we investigate a class of linearized equations. In section 4, we construct a class of differential 4 uniform functions by exchanging the two points in the value table of the Gold functions. Finally, some conclusions are given.

## 2 Preliminaries

Let  $F_{2^n}$  be the field with  $2^n$  elements, which is also a vector space over  $F_2$  with dimension  $n$ , and  $F_{2^n}^*$  denote by the set of all nonzero elements in  $F_{2^n}$ . The trace function from  $F_{2^n}$  to its subfield  $F_{2^m}$  is defined by

$$Tr_m^n(x) = \sum_{i=0}^{n/m-1} x^{2^{im}}.$$

A function  $f : F_{2^n} \rightarrow F_{2^n}$  can be represented as a univariate polynomial as follows:

$$f(x) = \sum_{i=0}^{2^n-1} a_i x^i, \quad a_i \in F_{2^n}.$$

If  $f(x)$  satisfies  $f(x)^2 = f(x)$  then  $f(x)$  is called a Boolean function over  $F_{2^n}$ . In this case,  $f(x)$  can be expressed as a sum of some trace functions [12]. The algebraic degree of  $f(x)$  is the maximal 2 weight of  $i$  such that  $a_i \neq 0$ .

Let us recall some notations on vector Boolean functions which are related to resistance to differential cryptanalysis [1].

**Definition 2.1.** Let  $F : F_{2^n} \rightarrow F_{2^n}$  be a function. The derivative of  $F(x)$  with respect to  $a$  is defined by

$$D_a F(x) = F(x + a) + F(x), \quad x \in F_{2^n}.$$

For any  $(a, b) \in F_{2^n}^* \times F_{2^n}$ , we denote by

$$\Delta_F(a, b) = \{x \in F_{2^n} \mid D_a F(x) = b\}, \quad \delta_F(a, b) = |\Delta_F(a, b)|,$$

where ‘ $|\cdot|$ ’ stands for the cardinality of a set. Then

$$\delta(F) = \max_{a \neq 0, b \in F_{2^n}} \delta_F(a, b) \geq 2$$

is called the differential uniformity of  $F(x)$ , and the functions for which equality holds are said to be almost perfect nonlinear (APN) functions.

For a vector Boolean function  $F(x)$  over  $F_{2^n}$ , Its Walsh transform is defined by

$$W_F(a, b) = \sum_{x \in F_{2^n}} (-1)^{Tr_n^1(aF(x)+bx)}, \quad (a, b) \in F_{2^n}^* \times F_{2^n}.$$



The nonlinearity of  $F(x)$  is defined as

$$n(F) = 2^{n-1} - \frac{1}{2} \max_{(a,b) \in F_{2^n} \times F_{2^n}} |W_F(a,b)|.$$

The nonlinearity of  $F(x)$  is a measure of vulnerability to linear cryptanalysis.

### 3 A Class of Linearized Equations

In this section we investigate solutions in  $F_{2^n}$  of the linearized equation  $ax^{2^r} + bx + c = 0, a, b, c \in F_{2^n}, a \neq 0$  for latter usage. First we give a preliminary lemma, which is the generalization of Lemma 7.19 in [20].

**Lemma 3.1.** Let  $\alpha \in F_{2^n}$ . For  $0 \neq r \in Z_n$ , the equation

$$x^{2^r} + x + \alpha = 0 \tag{1}$$

has solutions in  $F_{2^n}$  if and only if

$$Tr_e^n(\alpha) = 0 \tag{2}$$

where  $e = \gcd(n, r)$ . Moreover, if Eq.(2) holds, then Eq.(1) has  $2^e$  solutions in  $F_{2^n}$ .

**Proof.** It is clear that Eq.(2) holds if Eq.(1) has a solution in  $F_{2^n}$ . Conversely, assume that  $Tr_e^n(\alpha) = 0$ . Let  $n = ke, r = se$  and  $\gcd(k, s) = 1$ . It can be verified that for  $y \in F_{2^n}$

$$y + y^{2^r} + y^{2^{2r}} + \dots + y^{2^{(k-1)r}} = Tr_e^n(y). \tag{3}$$

Choose a  $\vartheta \in F_{2^n}$  such that  $Tr_e^n(\vartheta) = 1$ . Let

$$\beta = \alpha\vartheta^{2^r} + (\alpha + \alpha^{2^r})\vartheta^{2^{2r}} + \dots + \left( \sum_{i=0}^{k-2} \alpha^{2^{ir}} \right) \vartheta^{2^{(k-1)r}}.$$

By the formula (3) we have  $\beta^{2^r} + \beta = \alpha$ , that is,  $\beta$  is a solution of Eq.(1). Suppose  $\beta + \gamma$  is another solution of Eq.(1), then we have  $\gamma^{2^r} = \gamma$ . Which shows that  $\gamma \in F_{2^n} \cap F_{2^r} = F_{2^e}$ . Therefore, Eq.(1) has either no solution or  $2^e$  solutions in  $F_{2^n}$ .

**Theorem 3.2.** Let  $0 \neq r \in Z_n$  with  $\gcd(n, r) = e$  and

$$ax^{2^r} + bx + c = 0 \tag{4}$$

be a equation over  $F_{2^n}$ , where  $a, b, c \in F_{2^n}$  and  $a \neq 0$ . When  $b = 0$ , Eq.(4) has a unique solution  $x = (c/a)^{2^{n-r}}$ . When  $b \neq 0$ , Eq.(4) has solutions if and only if

$$a^{-1}b \in S, \quad \text{and} \quad Tr_e^n \left( \frac{c}{b\theta} \right) = 0 \tag{5}$$

here  $S = \{x^{2^e-1} \mid x \in F_{2^n}^*\}$ , and  $\theta \in F_{2^n}^*$  such that  $\theta^{2^r-1} = a^{-1}b$ . Moreover, if the conditions (5) holds then Eq.(4) has  $2^e$  solutions in  $F_{2^n}$ .

**Proof.** When  $b = 0$ , the result is clear. When  $b \neq 0$ , if Eq.(4) has a solution  $\beta \in F_{2^n}$ , for any  $y \in F_{2^n}^*$ , substituting  $x = y + \beta$  into Eq.(4) we get

$$ay^{2^r} + by = 0,$$

which implies that  $a^{-1}b \in \{y^{2^r-1} \mid y \in F_{2^n}^*\} = S$  because that  $\gcd(2^r - 1, 2^n - 1) = 1$ .

Let  $a^{-1}b = \theta^{2^r-1}$  for some  $\theta \in F_{2^n}$ , and  $x = z\theta$ . Then Eq.(4) is equivalent to

$$x^{2^r} + x + \frac{c}{b\theta} = 0 \tag{6}$$

By Lemma 3.1, we have  $Tr_e^n(c/(b\theta)) = 0$ .

Conversely, if the conditions (5) holds then  $a^{-1}b = \theta^{2^r-1}$  for some  $\theta \in F_{2^n}$ . By the linear transform  $x = z\theta$ , Eq.(4) is equivalent to Eq.(6). According to  $Tr_e^n(c/(b\theta)) = 0$  and Lemma 3.1, Eq.(6), and also Eq.(4) has  $2^e$  solutions in  $F_{2^n}$ .

**Corollary 3.3.** Let  $0 \neq r \in Z_n$  with  $\gcd(n, r) = 1$ . When  $b = 0$ , Eq.(4) has a unique solution  $x = (c/a)^{2^{n-r}}$ . When  $b \neq 0$ , Eq.(4) has

$$\begin{cases} \text{no solutions,} & \text{if } Tr_1^n \left( \frac{ca^{r(s)}}{b^{r(s)+1}} \right) = 1, \\ \text{two solutions,} & \text{if } Tr_1^n \left( \frac{ca^{r(s)}}{b^{r(s)+1}} \right) = 0 \end{cases}$$

where  $r(s) = \sum_{i=0}^{s-1} 2^{ir}$ ,  $s$  is the inverse of  $r$  in  $Z_n$ .

**Proof.** Since  $\gcd(r, n) = 1$ ,  $\gcd(2^r - 1, 2^n - 1) = 1$ , and so

$$(2^r - 1)r(s) = \sum_{i=0}^{s-1} 2^{ir} \equiv 1 \pmod{2^n - 1}.$$

There exists  $\theta \in F_{2^n}$  such that  $\theta^{2^r-1} = ba^{-1}$  and  $\theta = (ba^{-1})^{r(s)}$ . Let  $x = \theta y$ , then Eq.(4) is equivalent to

$$y^{2^r} + y + \frac{c}{b\theta} = 0 \tag{7}$$

By Lemma 3.1, Eq.(7), also Eq.(4) has solutions if and only if

$$Tr_1^n\left(\frac{c}{b\theta}\right) = Tr_1^n\left(\frac{ca^{r(s)}}{b^{r(s)+1}}\right) = 0 \tag{8}$$

Furthermore, according to Lemma 3.1 again, if Eq.(8) holds, then Eq.(4) has 2 solutions since  $\gcd(n, r) = 1$ .

**Remark 3.4.** When  $r = 1$ , then  $r(s) = 1$ , Corollary 3.3 is just theorem 7.20 in [20], and so it is a generalization of theorem 7.20 of [20].

### 4 A Class of 4 Differential Uniform Functions

As the first application of the results of the section 3 we determine all pairs  $(a, b) \in F_{2^n}^* \times F_{2^n}$  such that for two inputs with difference  $a$  the Gold function has the output  $b$ .

**Proposition 4.1.** Let  $F(x) = x^{2^r+1}$  be the Gold APN function over  $F_{2^n}$ , where  $\gcd(n, r) = 1$ . Then  $|\Delta_F(a, b)| = 2$  if and only if  $Tr_1^n(1 + b/a^{2^r+1}) = 0$ , here ‘ $|\cdot|$ ’ stands for the cardinality of a set.

**Proof.** For  $(a, b) \in F_{2^n}^* \times F_{2^n}$ , let

$$(x + a)^{2^r+1} + x^{2^r+1} = b,$$

which is equivalent to

$$ax^{2^r} + a^{2^r}x + a^{2^r+1} + b = 0.$$

By Corollary 3.3,  $|\Delta_F(a, b)| = 2$  if and only if  $Tr_1^n(1 + b/a^{2^r+1}) = 0$ .

Next we investigate the differential uniformity, algebraic degree and nonlinearity of the functions constructed from Gold APN functions by exchanging two values of the functions. First we recall the following general result.

**Lemma 4.2[19].** Let  $F(x)$  be a function from  $F_{2^n}$  to  $F_{2^n}$ . Let  $G(x)$  be the function constructed by change two values of  $F(x)$ , i.e. for any  $u \neq v \in F_{2^n}$ ,

$$G(x) = \begin{cases} F(v), & x = u, \\ F(u), & x = v, \\ F(x), & x \neq u, v. \end{cases}$$

Then  $\Delta_G(a, b) \subseteq \Delta_F(a, b) \cup \{u, u + a, v, v + a\}$  for  $(a, b) \in F_{2^n}^* \times F_{2^n}$ , and  $\delta(F) - 4 \leq \delta(G) \leq \delta(F) + 4$  and  $n(F) - 2 \leq n(G) \leq n(F) + 2$ . Moreover, if  $\delta(F) = 2$  then  $\delta(F) \in \{2, 4\}$ .

**Lemma 4.2.** Let  $F(x) = x^{2^r+1}$ ,  $\gcd(n, r) = 1$ , be the Gold APN function over  $F_{2^n}$ . If  $n$  is odd then the nonlinearity of  $F(x)$  is  $2^{n-1} - 2^{(n-1)/2}$ ; If  $n$  is even then the nonlinearity of  $F(x)$  is  $2^{n-1} - 2^{n/2}$ .

**Proof.** The case of odd  $n$  has been proved in [11]. The proof of the even case is similar to the odd case, and is included here for convenience.

First we compute the value of Walsh transform at some  $(u, v) \in F_{2^n}^* \times F_{2^n}$ ,

$$\begin{aligned} W_F(u, v)^2 &= \sum_{x, y \in F_{2^n}} (-1)^{Tr_1^n\left(ux^{2^r+1} + u(x+y)^{2^r+1} + vx + v(x+y)\right)} \\ &= \sum_{y \in F_{2^n}} (-1)^{Tr_1^n\left(uy^{2^r+1} + vy\right)} \sum_{x \in F_{2^n}} (-1)^{Tr_1^n\left(u(y^{2^r} + y^{2^r})x\right)} \\ &= 2^n \sum_{y \in S} (-1)^{Tr_1^n\left(uy^{2^r+1} + vy\right)}, \end{aligned}$$

here  $S = \{y \in F_{2^n} \mid y^{2^r} + y^{2^{-r}} = 0\}$ . Since  $\gcd(2r, n) = 2$ , for any  $y \in F_{2^n}$ , it satisfies  $y^{2^r} + y^{2^{-r}} = 0$  if and only if  $y \in F_{2^n} \cap F_{2^{2r}} = F_4$ , and so  $S = F_4$ . It is clear that  $3 \mid (2^r + 1)$  since  $r$  is odd. Then

$$W_F(u, v)^2 = 2^n (-1)^{Tr_1^n(u)} \sum_{y \in F_4} (-1)^{Tr_1^n(vy)},$$

and so

$$n(F) = 2^{n-1} - \frac{1}{2} \max_{(u, v) \in F_{2^n}^* \times F_{2^n}} |W_F(u, v)| = 2^{n-1} - 2^{n/2}.$$

Next we prove our main results.

**Proposition 4.3.** Let  $F(x) = x^{2^r+1}$ ,  $\gcd(n, r) = 1$ , be the Gold APN function over  $F_{2^n}$ . Let  $\beta$  be a primitive element of  $F_{2^n}$ . For any  $i \in \{0, 1, 2, \dots, 2^n - 2\}$ , define

$$F_{0i}(x) = \begin{cases} \beta^{i(2^r+1)}, & x = 0, \\ 0, & x = \beta^i, \\ x^{2^r+1}, & x \neq 0, \beta^i. \end{cases}$$

Then we have

$$F_{0i}(x) = \beta^{i(2^r+1)} + \beta^{i(2^r+1)} \sum_{k=1}^{2^n-3} \left( \beta^{ik} x^{2^n-1-k} \right)^{2^r+1},$$

and the differential uniform  $\delta(F_{0i}) = 4$ , and the algebraic degree of  $F_{0i}(x)$  is  $n-2$ , and the nonlinearity of  $F_{0i}(x)$  is  $2^{n-1} - 2^{(n-1)/2} - 2 \leq n(F_{0i}) \leq 2^{n-1} - 2^{(n-1)/2} + 2$  if  $n$  is odd, and  $2^{n-1} - 2^{n/2} - 2 \leq n(F_{0i}) \leq 2^{n-1} - 2^{n/2}$  if  $n$  is even.

**Proof.** It is easy to verify that

$$F_{0i}(x) = \beta^{i(2^r+1)} + \beta^{i(2^r+1)} \sum_{k=1}^{2^n-3} \left( \beta^{ik} x^{2^n-1-k} \right)^{2^r+1}.$$

To determine the algebraic degree of  $F_{0i}(x)$ , One can check that  $(2^n - 1 - k)(2^r + 1) \bmod 2^n$  has the maximal 2 weight  $n - 2$  for  $1 \leq k \leq 2^n - 3$ . Hence, the algebraic degree  $F_{0i}(x)$  is  $n - 2$ . According to Lemma 4.2 and Lemma 4.3, it is clear that the nonlinearity of  $F_{0i}(x)$  is  $2^{n-1} - 2^{(n-1)/2} - 2 \leq n(F_{0i}) \leq 2^{n-1} - 2^{(n-1)/2} + 2$  if  $n$  is odd, and  $2^{n-1} - 2^{n/2} - 2 \leq n(F_{0i}) \leq 2^{n-1} - 2^{n/2}$  if  $n$  is even.

Next we determine the differential uniformity of  $F_{0i}(x)$ . For  $\beta^i \neq a \in F_{2^n}$ , we have

$$D_a F_{0i}(x) = \begin{cases} a^{2^r+1} + \beta^{i(2^r+1)}, & x = 0, a, \\ (a + \beta^i)^{2^r+1}, & x = \beta^i, \beta^i + a, \\ (x + a)^{2^r+1} + x^{2^r+1}, & x \notin \{0, a, \beta^i, \beta^i + a\}. \end{cases}$$

Set  $(x + a)^{2^r+1} + x^{2^r+1} = a^{2^r+1} + \beta^{i(2^r+1)}$ , which is equivalent to

$$ax^{2^r} + a^{2^r}x + \beta^{i(2^r+1)} = 0 \tag{9}$$

By Corollary 3.3, it has solutions in  $F_{2^n}$  if and only if  $Tr_1^n \left( (\beta^i a^{-1})^{2^r+1} \right) = 0$ . Let  $a_0 \in F_{2^n}$  satisfies this equation, and Eq.(9) has the corresponding two solutions  $\{x_0, x_0 + a_0\}$ .

Let  $b = a_0^{2^r+1} + \beta^{i(2^r+1)}$ . we have

$$\Delta_{F_{0i}}(a_0, b) = \{0, a_0, x_0, x_0 + a_0\}$$

Therefore,  $\delta(F_{0i}) = 4$ .

**Proposition 4.4.** Let  $F(x) = x^{2^r+1}$ ,  $\gcd(n,r) = 1$ , be the Gold APN function over  $F_{2^n}$ . Let  $\beta$  be a primitive element of  $F_{2^n}$ . For any  $i \neq j \in \{0, 1, 2, \dots, 2^n - 2\}$ , define

$$F_{ij}(x) = \begin{cases} \beta^{j(2^r+1)}, & x = \beta^i, \\ \beta^{i(2^r+1)}, & x = \beta^j, \\ x^{2^r+1}, & x \neq \beta^i, \beta^j. \end{cases}$$

Then  $\delta(F_{ij}) = 4$ , and

$$F_{ij}(x) = x^{2^r+1} + (\beta^{i(2^r+1)} + \beta^{j(2^r+1)}) \left( \prod_{c \in F_{2^n}, c \neq \beta^i} \frac{x-c}{\beta^i - c} + \prod_{c \in F_{2^n}, c \neq \beta^j} \frac{x-c}{\beta^j - c} \right) \tag{10}$$

and the algebraic degree of  $F_{ij}(x)$  is  $n-1$ , and the nonlinearity of  $F_{ij}(x)$  is the same as that of  $F_{0i}(x)$ .

**Proof.** By Lagrange interpolation formula, it is easy to verify that  $F_{ij}(x)$  has the explicit expression as (10). Since

$$\prod_{c \in F_{2^n}, c \neq \beta^i} \frac{x-c}{\beta^i - c} = \prod_{\alpha \in F_{2^n}^*} \alpha \left( \prod_{c \in F_{2^n}, c \neq \beta^i} (x-c) \right),$$

we can check that the coefficient of  $x^{2^n-1}$  in  $F_{ij}(x)$  is zero, and the coefficient of  $x^{2^n-2}$  in  $F_{ij}(x)$  is nonzero, and so the algebraic degree of  $F_{ij}(x)$  is  $n-1$ . It is clear that the nonlinearity is the same as that of  $F_{0i}(x)$  by Lemma 4.2 and Lemma 4.3.

Next we compute the differential uniformity of  $F_{ij}(x)$ . For

$\beta^i + \beta^j \neq a \in F_{2^n}^*$ , we have

$$D_a F_{ij}(x) = \begin{cases} (\beta^i + a)^{2^r+1} + \beta^{j(2^r+1)}, & x = \beta^i, \beta^i + a, \\ (\beta^j + a)^{2^r+1} + \beta^{i(2^r+1)}, & x = \beta^j, \beta^j + a, \\ (x+a)^{2^r+1} + x^{2^r+1}, & x \notin \{\beta^i, \beta^i + a, \beta^j, \beta^j + a\}. \end{cases} \quad \text{Set}$$

$(x+a)^{2^r+1} + x^{2^r+1} = (\beta^i + a)^{2^r+1} + \beta^{j(2^r+1)}$ , which is equivalent to

$$ax^{2^r} + a^{2^r}x + \beta^{i2^r}a + \beta^i a^{2^r} + \beta^{i(2^r+1)} + \beta^{j(2^r+1)} = 0 \tag{11}$$

By Corollary 3.3, Eq.(11) has solutions in  $F_{2^n}$  if and only if

$$Tr_1^n \left( (\beta^{i(2^r+1)} + \beta^{j(2^r+1)}) a^{-(2^r+1)} \right) = 0 \tag{12}$$

It is clear that there exists  $a' \in F_{2^n}$  such that Eq.(12) holds.

Denote by  $y_0, y_0 + a'$  two solutions of the corresponding Eq.(11). Let  $b = (\beta^i + a')^{2^r+1} + \beta^{j(2^r+1)}$ , and then  $\Delta_{F_{ij}}(a', b) = \{\beta^i, \beta^j + a', y_0, y_0 + a'\}$ . So  $\delta(F_{ij}) = 4$ .

## 5 Conclusions

In this paper we determine the number of solutions of a class of linearized equations, and use the related results to prove that the differential uniformity of the functions constructed by exchanging two points in the value table of the Gold functions is 4, and the derived functions have high algebraic degree and high nonlinearity.

**Acknowledgments.** The work is supported by the National Science Foundation of China under Grant 11101311 and the Youth Foundation of Mathematical Tianyuan of China under Grant 10726020.

## References

1. Berger, T.P., Canteaut, A., Charpin, P., Chapuy, Y.L.: On almost perfect nonlinear functions over. *IEEE Trans. Inform. Theory* 52(9), 4160–4170 (2006)
2. Beth, T., Ding, C.: On Almost Perfect Nonlinear Permutations. In: Helleseht, T. (ed.) *EUROCRYPT 1993*. LNCS, vol. 765, pp. 65–76. Springer, Heidelberg (1994)
3. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. *J. Cryptology* 4(1), 3–72 (1991)
4. Bracken, C., Byrne, E., Markin, N., McGuire, G.: New families of quadratic almost perfect nonlinear trinomials and multinomials. *Finite Field Appl.* 14(3), 703–714 (2008)
5. Bracken, C., Leander, G.: A highly nonlinear differential 4 uniform power mapping that permutes fields of even degree. *Finite Fields Appl.* 16(4), 231–242 (2010)
6. Budaghyan, L., Carlet, C., Felke, P., Leander, G.: An infinite class of quadratic APN functions which are not equivalent to power mappings. In: *IEEE ISIT 2006*, Seattle, USA (2006)
7. Budaghyan, L., Carlet, C.: Classes of quadratic APN trinomials and hexanomials and related structures. *IEEE Trans. Inform. Theory* 54(5), 2354–2357 (2008)
8. Budaghyan, L., Carlet, C., Felke, P., Leander, G.: Construction new APN functions from known ones. *Finite Field Appl.* 15(2), 150–159 (2009)
9. Carlet, C.: Boolean Functions for Cryptography and Error Correcting Codes. In: Crama, Y., Hammer, P. (eds.) *The Monography Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, pp. 257–397. Cambridge Univ. Press (2010)
10. Dobbertin, H.: Almost perfect nonlinear functions over  $GF(2^n)$ : a new case for  $n$  divisible by 5. In: *Proceeding of 5th Finite Field and their Applications*, pp. 113–121 (2000)
11. Gold, R.: Maximal recursive sequences with 3-valued recursive cross-correlation functions. *IEEE Trans. Inform. Theory* 14(1), 154–156 (1968)
12. Golomb, S.W., Gong, G. (eds.): *Signal Designs With Good Correlation: For Wireless Communications, Cryptography and Radar Applications*. Cambridge University Press, Cambridge (2005)

13. Kasami, T.: The weight enumerators for several classes of subcodes of the second order binary Reed-Muller codes. *Inform. Control* 18(4), 369–394 (1971)
14. Knudsen, L.: Truncated and Higher Order Differentials. In: Preneel, B. (ed.) *FSE 1994*. LNCS, vol. 1008, pp. 196–211. Springer, Heidelberg (1995)
15. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: Helleseht, T. (ed.) *EUROCRYPT 1993*. LNCS, vol. 765, pp. 386–397. Springer, Heidelberg (1994)
16. Nyberg, K.: Differentially Uniform Mappings for Cryptography. In: Helleseht, T. (ed.) *EUROCRYPT 1993*. LNCS, vol. 765, pp. 55–64. Springer, Heidelberg (1994)
17. Niho, Y.: Multi-valued cross-correlation functions between two maximal linear recursive sequences. Ph D. thesis, University of South California (1972)
18. Nakagawa, N., Yoshiara, S.: A Construction of Differentially 4-Uniform Functions from Commutative Semifields of Characteristic 2. In: Carlet, C., Sunar, B. (eds.) *WAIPI 2007*. LNCS, vol. 4547, pp. 134–146. Springer, Heidelberg (2007)
19. Yu, Y., Wang, M., Li, Y.: A new method for constructing differential 4-uniform permutations from known ones, <http://eprint.iacr.org/2011/047>
20. Wan, Z.: *Lectures on Finite Fields and Galois Rings*. World Scientific Pub. Co. Inc. (2003)



# The Stability of USO to Mars Orbiter Influence

Zhen Wang<sup>1,2,3,\*</sup>, Kun Shang<sup>4</sup>, Nan Ye<sup>5</sup>, Na Wang<sup>1</sup>, and Jin Song Ping<sup>4</sup>

<sup>1</sup> Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, P.R. China

<sup>2</sup> Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, P.R. China

<sup>3</sup> Shanghai key laboratory of Space Navigation and Position

<sup>4</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

<sup>5</sup> State Key Laboratory of Space Dynamics, Xian 710043, China  
wangzh@xao.ac.cn, shang.34@osu.edu, nl\_1024@163.com,  
na.wang@uao.ac.cn, pjs@shao.ac.cn

**Abstract.** The China's first Mars orbiter "YingHuo-1" (YH-1) will be launched in the near future. To solve the tracking problem of Mars satellite, the kernel payload of Ultra Stable Oscillator (USO) of "YingHuo-1" is used for frequency standard, and the transmitter generates a standard downlink radio signal. The ground astronomical Very Long Baseline Interferometry (VLBI) system will be used to receive the radio signal. In this article, to investigate the mechanism of the tracking accuracy of the orbit determination of YH-1, the experimental data of YingHuo-1's USO was processed, using Solar-system Model and Orbit Determination (SMOD) software system developed at Xinjing Astronomical Observatory. The real-time precise orbit determination can be supported by the simulation experimental result, when the frequency stability of USO is about 10-12.

**Keywords:** YingHuo-1, Ultra Stable Oscillator, frequency standard.

## 1 Introduction

Mars is the fourth planet, who is a terrestrial planet, in the Solar System. The distance is about  $3.6 \times 10^8$  km from the earth to the Mars. There are the neutral atmospheres and ionospheres, the valleys, the volcanoes, the deserts, and so on in Mars. Since the 20th century 60's, the United States, the Soviet Union and the European Space Agency launched a large number of Mars orbiters, the orbital probes were carried out on the Mars comprehensive, multi-detection means, and accumulating amounts of observational data. China's first Mars satellite of YH-1[1,2], that mission will mainly

---

\* Weng zhen: Xinjiang Astronomical Observatory,  
No.150 South Science Road, Urumqi, Xinjiang 830011, P.R. China  
Email: wangzh@xao.ac.cn

investigate the atmosphere and ionosphere of the Mars. In view of absence of any uplink system in the China's Network (Shanghai, Urumqi, Beijing and Kunming), the X-band's receiver and transmitter system have been adopted for onboard communication. Different from common deep space mission, the open loop tracking method will be applied for the tracking and navigation requirement of YH-1 mission, like as One-way Doppler, Differenced One-way Doppler (DOD), and Seam Beam VLBI[3], and so on.

## 2 Experiment Description

YH-1's USO experiments were conducted in March 2010. Test equipment connection diagram was shown in Fig.1. Radio transmitter of YH-1 was used during the series of experiments to generate the single-frequency at a frequency of 8 424.407 040 MHz. Radio transmitter of YH-1 and power attenuator were placed in constant temperature room, by 150 m outdoor cable with S/X dual-frequency refrigeration receiver connection, the radio signal was received by digital terminal recording device of K5[4].

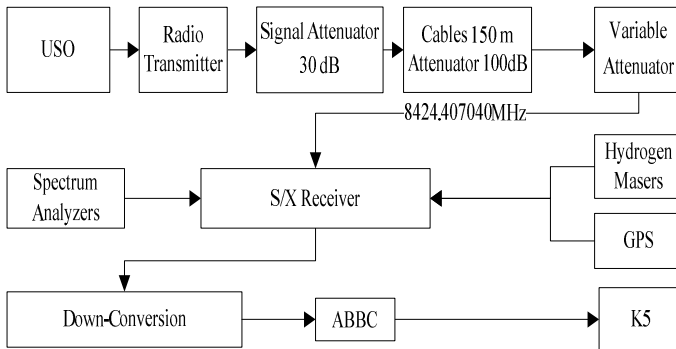


Fig. 1. USO equipment connection diagram

## 3 Data Analysis

Doppler frequency data of March 2010 were extracted from K5 in this article. First of all, the S/X receiver received the RF signals of YH-1. Then, the signals were reduced the lower bandwidth of frequency by the digital filter. Finally, the signals were taken by K5. Be relative to the data of USO was about 75 h, the frequency changes were shown in Fig.2. After starting the transmitter, it need some time to stabilize the equipment. The frequency drift was about 3 Hz, before stabilizing the frequency of USO. The some obvious bad data had been removed. The frequency of USO was changed within a range of about 0.6 Hz.

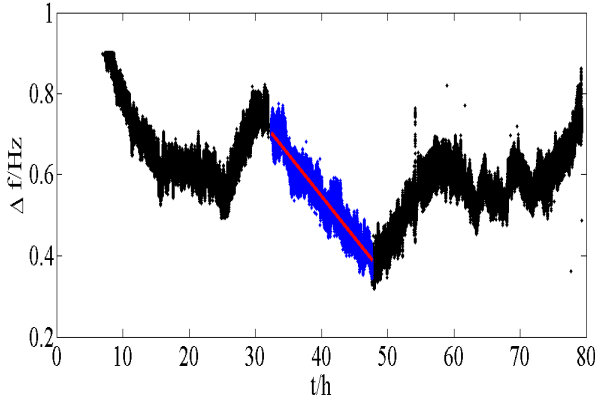


Fig. 2. Trend of Frequency

### 4 Precision Analysis

The important indicators of frequency standard are frequency accuracy, frequency stability and frequency drift rate. The standard of frequency is to provide high reliability and high precision frequency signal by oscillator circuit, such as USO. It is similar to the heart of the system of equipment, in order to maintain reunification of time in the system. High reliability indicates that the system error and random error are smaller, high precision show that the data of dispersion is small. Reducing the system error of the frequency is to improve the reliability of the frequency. Reducing the discrete level of frequencies is to improve the frequency stability. A good performance standard frequency should be accurate and sure.

In the main measurement error is the frequency deviation of USO. The simple frequency evolution is given by

$$f_T = f_{T_0} + Bt + A \cdot rand(t) \tag{1}$$

In the formula  $f_{T_0}$  is the nominal frequency value. B is linear evolution factor. A is the noise intensity which is related to Allen variance [5] (AVAR).

In the one-way Doppler measurement, the observable accuracy depends on the accuracy of USO. The relationship between measurement systematic error and the USO accuracy is given by

$$\Delta f_R^{1-way} = \Delta f_T \left( 1 - \frac{v}{c} \right) \approx \Delta f_T \tag{2}$$

The DOD measurement error is given by

$$\Delta D - f_{R_{12}}^{1-way} \approx \Delta f_T(t_{T_2}) - \Delta f_T(t_{T_1}) \tag{3}$$

Where  $t_{T1}$  and  $t_{T2}$  are related to the different emission moment.

From what have been discussed above, we can get these conclusions. The systematic accuracy of One-way Doppler is equal to the accuracy of the USO on satellite. The DOD accuracy depends on the frequency change between two emission

moments  $t_{TI}$  and  $t_{TI}$ . So non-differential measurement accuracy depends on the frequency source accuracy, and differential measurement accuracy depends on short time stability of the frequency source. Simulation result shows that time difference of two transmission moments is about 0.01s, the frequency change is under 1 mHz. On the other hand, the accuracy of source includes long term drift which can reach several Hz. So the accuracy of differential result is higher than non-differential result.

The AVAR is the most common time domain measure of frequency stability. The original non-overlapped AVAR (two-sample variance), is the standard time domain measure of frequency stability. It is defined as

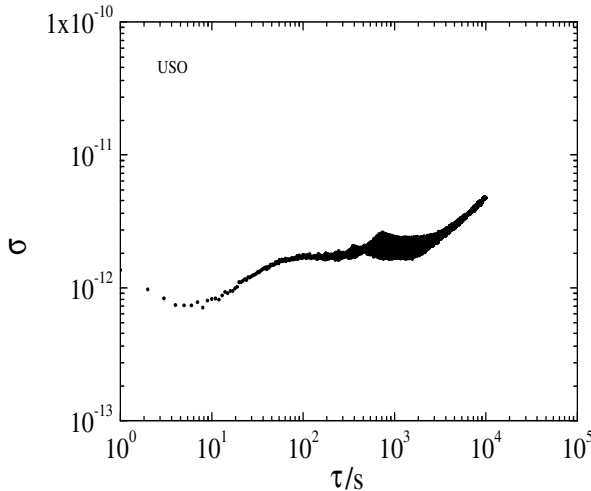
$$\sigma_y^2(\tau) = \frac{1}{2} \left\langle \left( \overline{y_{n+1}} - \overline{y_n} \right)^2 \right\rangle \tag{4}$$

Where  $y_n$  is the  $n$  th of data, the brackets  $\langle \rangle$  denote the expectation value,  $\tau$  is the interval time.

In March 2010, the AVAR of USO is showed in Fig.3 and Tab.1, respectively.

**Table 1.** AVAR of USO

times(s)	1	10	$10^2$	$10^3$	$10^4$
AVAR	$5.3e^{-13}$	$6.4e^{-13}$	$0.9e^{-12}$	$1.3e^{-12}$	$4.6e^{-12}$



**Fig. 3.** AVAR of USO

## 5 Influence of OD of USO

The Mars orbiter can be observed, the main subject to the Earth's rotation visual angle and low inclination orbit. The single station visual arc is only about 10 h. The China VLBI Network antennas only are distributed in the domestic local, to improve the visual arcs contribution is less. Therefore, the irregular change of USO will not affect

the all observation data. But sub-arc length observation data of about 10 h can be affected. In the orbit determination (OD) strategy of YH-1, traditional method of sub-arc frequency parameters is still used.

In the simulate experiment, the irregular change of USO is mainly caused by the change of external temperature. It's change cycle is about 24 h. YH-1 is high orbital satellite, and the sun's shadow is rarely. The change of the temperature effect depends on the attitude and the position of orbiter. So, the USO signal change cycle is mainly orbital cycle of about 72 h. In the OD, the frequency of USO is used quadratic polynomial, and the frequency shift and drift are estimated.

## 6 The Simulation Conditions

Interplanetary orbit determination is one of the key issues associated with deep space exploration. Precise orbit determination (POD) ensures that interplanetary vehicles can be controlled along pre-planned spaceflights, and is a prerequisite for data analysis and scientific research. The orbit determination software system for Mars satellites has been recently developed at Xinjing Astronomical Observatory, the basic institutional objective being to meet Chinese Mars exploration project requirements. The software uses dynamic statistical orbit determination, and is mainly divided into three modules: (1) Numerical integration of orbit and state transition matrix; (2) Measure modeling and related derivatives; (3) Statistical estimation of normal equations using least squares estimation. Currently, the software can use any of the following measurement modes: One-way Doppler, One-way Ranging, Two-way Doppler, Two-way Ranging, Three-way Doppler, and Three-way Ranging.

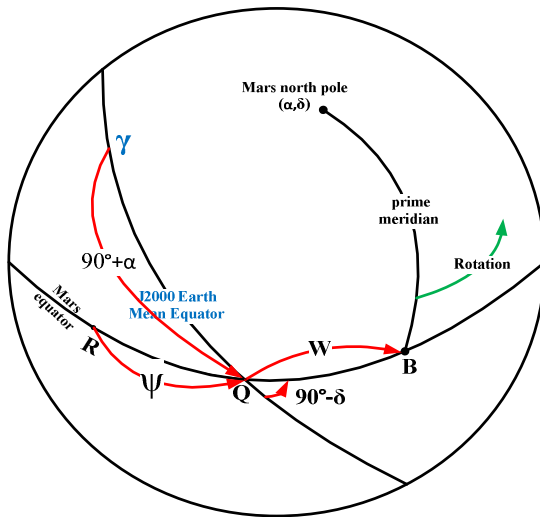


Fig. 4. IAU Mars orientation model

The Mars inertial coordinate system[6] for the orbit of YH-1 is shown in Fig4. Near Mars point of the YH-1's orbit is about 800 km, far Mars point is about 80 000

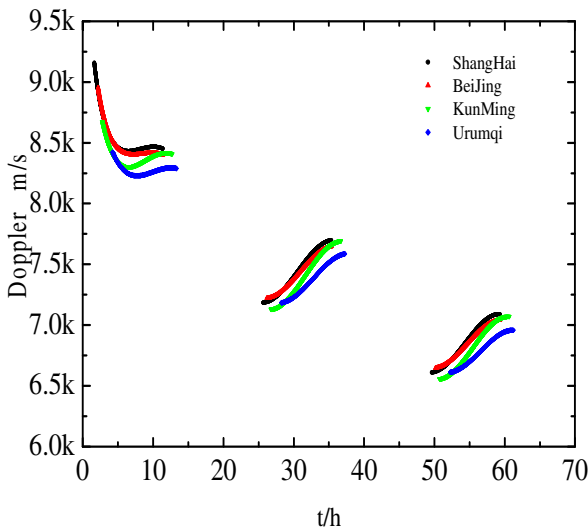
km, orbital inclination is about 5 degrees, the other orbital parameters are 0, and start time of simulation is Sep.1, 2012 UTC 00:00:00, under the cartesian coordinates of the International Celestial Reference Frame, as showed Tab.2:

**Table 2.** Position and Velocity of initial YH-1

	x	y	z
Position(km)	2825.821839	3102.054392	0
Velocity(km/sec)	-2.417307	2.202050	2.956802

The error of frequency measurement is equivalent to  $\Delta f = f_0 \cdot \sigma$ , the error of radial velocity is equivalent to about  $\Delta v = c \cdot \sigma$ . For the transmitting frequency of the X-band, the error caused by about 8.4 mHz, conversion to the noise caused by one-way speed was about 0.3 mm/s. The noise of simulation data presented the normal distribution, mean was 0, standard deviation were taken 0.3 mm/s and 3 mm/s, respectively. Frequency drift estimated as shown in Fig.2, in the experiment, and the stabile frequency of change of was used a blue line, red line for fitting results. Simulation data was used two ways, with frequency drift and without frequency drift, respectively.

For observed data of One-way doppler in four stations of China’s VLBI Network, the observation arc of simulation of a circle time was 72 h. The circle was divided into three segments of visual arc, as shown in Fig. 5.



**Fig. 5.** Visual arc of four stations

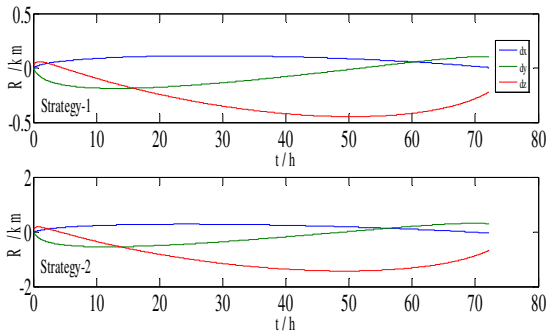
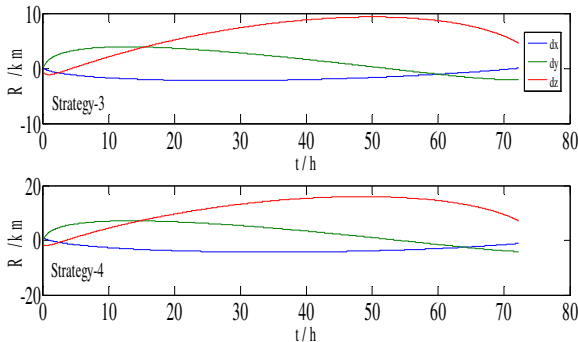
The solutions are the six kinds of orbit parameters, frequency offset and frequency drift. All of the simulation strategies were shown in Tab.3.

**Table 3.** OD simulation strategy

simulation strategy	noise (mm/s)	frequency offset/drift	orbit error (km)
1	0.3	without estimate	0.5
2	0.3	estimate	1.5
3	3	without estimate	10
4	3	estimate	15

## 7 The Simulation Results

The simulation results are as shown in Fig.6, Fig.7 corresponding four kinds of simulation strategy, respectively. The figure is displayed on the OD tracks with simulation orbit one order difference. When observed data noise is 0.3 mm/s, the error of the orbit of four stations is less than 500 m in three days. Adding the frequency within offset and drift, the orbit error increases to 1.5 km or so. When the observed data noise is 3 mm/s, the error of the orbit of four stations is less than 10 km in three days. With the frequency offset and drift, the orbit error increases to 15 km or so.

**Fig. 6.** Visual arc of four stations**Fig. 7.** Visual arc of four stations

## 8 Conclusions

The simulate accuracy error of satellite orbits is about 15 km, using the software system of Solar-system Model and Orbit Determination developed at Xinjing Astronomical Observatory, when the frequency stabilization of Ultra Stable Oscillator of YH-1 is  $10^{-12}$ , and the observed data noise is the 3 mm/s, taking into account the doppler frequency of the offset and the drift. The space weather of planet in the solar system can be explored, the deep space communication and navigation techniques can be tested, and the neutral atmosphere, ionosphere in radio occultation measurement, gravity field model of planet can be researched by these results.

**Acknowledgments.** The authors would like to thank Shanghai Astronomical Observatory and Xinjiang Astronomical Observatory.

Supported by Science & Technology Commission of Shanghai Municipality (06DZ22101), Shanghai Astronomical Observatory (Grant No.201005).

## References

1. Ping, J.-S., Shang, K.: Open loop tracking for Yinghuo-1 Mars Orbiter. *Physics* 38(11), 779–784 (2009) (in Chinese)
2. Shang, K., Jian, N.-C.: Open loop Doppler technology for the Mars Express tracking experiment. *Physics* 38(11), 799–803 (2009) (in Chinese)
3. Wang, Z., Ye, N., Shang, K.: The Contribution of Urumqi Astronomical Observatory in Lunar Exploration with Same Beam Interferometry. *Progress in Astronomy* 28(4), 424–432 (2010)
4. Jian, N.-C., Shang, K.: A digital open-loop Doppler processing prototype for deep-space navigation. *Sci. China Ser. G-Phys. Mech. Astron.* 52(12), 1849–1857 (2009)
5. Allan, D.W.: Statistics of Atomic Frequency Standard. In: *Proc. IEEE*, 54, 221–230 (1966)
6. Seidelmann, P.K.: *CeMDA*, vol. 98, pp. 155–180 (2007)



# Review of the Check Method in Metrology Support for Equipment

Wei Jiang

Department of Instrument Science and Technology, College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan, P.R. China, 410073  
vivian\_nudt@163.com

**Abstract.** Control of the quality of test and calibration process by checking is one of the important aspects in equipment measurement support process. To ensure the quality of measuring result, appropriate check method should be select on the basis of the requirement of measurement support assignment and different check objects. In this paper, check methods are summarized systematically, check purpose are presented, as well as the meanings and implementation procedure of check methods.

**Keywords:** Check Method, Check Purpose, Check Implementation, Measurement Equipment.

## 1 Introduction

In ISO/IEC17025-2005 "General Requirements of the Testing and Calibration Laboratories" and CNAL/AO01-2005 "Testing and Calibration Laboratory Accreditation Criteria", the period verification expressed as [1]: Implementation of verification for reference standards, benchmarks, pass standards or working standards and the standard material (reference material) in accordance with the defined procedures and schedule, in order to maintain the confidence of its calibration status. According to this statement we can understand the meaning of verification, first, for verification time, it should meet requirements of the schedule and procedures, which can be interpreted as self-imposed by legal metrology institutes; second, for verification object, it should be understood as the measurement standards and reference materials; third, the meaning and purpose of check is to maintain credibility of test or calibration status, to control quality of its process. From its meaning we can see, verification has an important position in the equipment measures support. The methods of verification discussed in this paper fit for military and civilian test lab, calibration laboratories, and other military and civilian metrology agencies using verification methods for quality control.

## 2 Check Purpose and Method

In the process of equipment measuring support, verification objects mainly include measurement, measuring equipment and measurement results. Appropriate verification

methods include statistical control verification methods for measurement process, verification methods for measuring equipment and for results, verification methods for each type of object further includes several specific implementation methods.

By using statistical verification methods of measurement process we can know the changes of measure standard in the measurement process, achieve continuous and long-term quality control of measurement process. Statistical method is used to control some key measurement process and the measurement process that is easy to loss control, as the measurement institutions of military is concerned, such as the highest standard or testing equipment for war field, the battlefield environment is always very complex, the use of these equipment for the measurement process is better to be controlled, to ensure that the equipment is always in good controlled condition to ensure accurate transmission of value, equipment performance be of reliable protection. Control charts[2] is the core tool for the statistical control of measurement process, control charts is a kind of map designed by statistical method which monitor process if it is in control state with determination, record, evaluation of process quality characteristic.

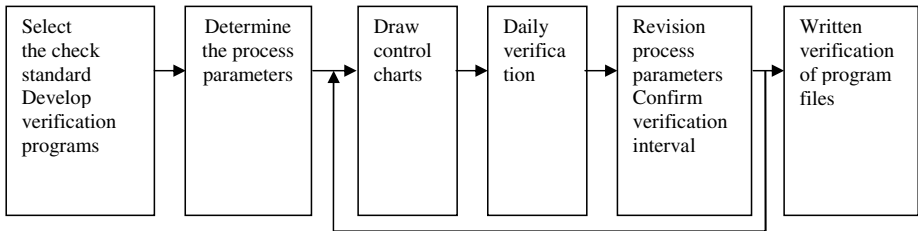


Fig. 1. Flow for measurement process statistical control verification

Verification of measuring equipment is necessary. In the current system of the amount of mass, the measuring equipment is periodically test, within the test cycle, it is often difficult to know whether the criteria have unexpected occurrence of inaccurate, the risk of the play of equipment performance due to inaccurate of measuring equipment are difficult to control. With verification of measuring equipment, including measurement systems, measurement standards and test equipment, we can detect changes, correct errors, to realize effective control of the measurement equipment within the qualifying period and to ensure the quality of measurement results, and thus play a security role for the play of equipment effectiveness.

The purpose of verification of measurement results is to verify the credibility of calibration, test and test results. In many cases it is necessary to verificate the measure results, such as: 1) those key measurements which have a higher accuracy requirements; 2) When the measuring equipment leave from the laboratory direct control (such as motor insurance or return from borrowed when need to check); 3) When the measurement results that is doubted or there is dispute to arbitration; 4) other, for whatever reason, it is necessary to verify the results of the measurement.

### 3 Implementation of the Check Method

No matter what kind of verification methods, its implementation contains several elements: selection of verification criteria, the application of verification, data recording, analysis and processing, show the assessment results. Among them, getting the assessment results is the ultimate goal; different methods have different analysis and assessment criteria.

#### 3.1 Measuring Process Statistical Control Verification Method

Method of measuring process statistical control verification refers to long-term statistical control of measuring process with verification of standards, it is the integrated control of measuring equipment, personnel, environmental conditions and other factors, it can be used to ensure the validity of the measurement process and measurement standards, so as to ensure the quality of measurement results. The process includes six parts [3] which is shown in Figure 1.

Verification standards are used to collect control database of measurement process, and measuring equipment, products or other objects which is measured by the process. Verification standards is the measured object of process to be controlled (or standard units), it would be measuring device, a product of this process or other objects measured by this process. Technical requirements for verification standard is they should coverage with parameters and values required to verify, values have good stability, regulatory requirements are they are easy to control by the agency to meet normal verification or random verification when measurement is doubtful.

Process parameters generally including the initial mean, standard deviation of the process experiment  $s_c$ , combined standard deviation  $s_p$ . To determine the process parameters, measure can be done at any time, but it is better to reflect the impact of various parameters as far as possible. The number of measurement group and repetitions of each group for the verification standard is determined by the actual situation. Assumed for  $m$  sets of measurements, each set repeat  $n$  times. Various statistical parameters calculated as follows:

Average of  $j$ -th set of measurements:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (1)$$

Where,  $x_{ij}$  is the  $i$ -th measured value of the  $j$ -th set of measurements,  $i = 1, 2, \dots, n$ .

The process mean is:

$$\bar{x} = \frac{1}{m} \sum_{j=1}^m \bar{x}_j \quad (2)$$

Experimental standard deviation of the process, that is, between subjects standard deviation:

$$s_c = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (\bar{x}_j - \bar{x})^2} \tag{3}$$

Experimental standard deviation of the j-set measurements, that is, experimental standard deviation for the j-group:

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \tag{4}$$

Combined standard deviation, that is, the statistical average of the within-subject standard deviation, as follows:

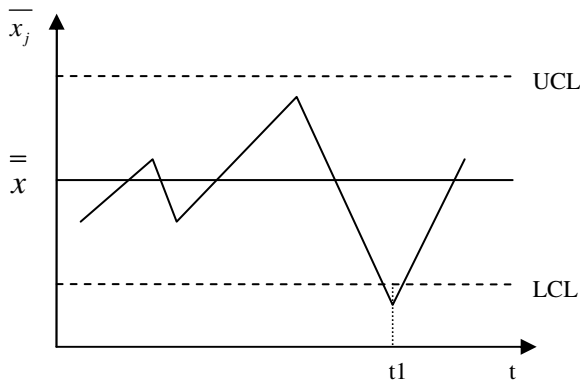
$$s_p = \sqrt{\frac{1}{m} \sum_{j=1}^m s_j^2} \tag{5}$$

For measurement control charts, the mean chart (X chart), range control chart (R chart) and standard deviation control chart (S chart) are recommended [4]. Mean control chart (X chart) is used to observe changes of the average of measurements, show long-term changes or those changes impacted by system of measurement process. Range control chart (R chart) is mainly used to observe the dispersion of measurement results, that is the fluctuations in the measurement results. Extreme difference is difference between the maximum and minimum of each measured value, expressed as R. When the number of observed is less, range control charts is more effective and convenient. Standard deviation control chart (S diagram) is mainly used to observe the dispersion of random errors; the standard deviation can be calculated with finite data column gained by each verification, expressed as S, it is more applicable to the case in which the number is much more. Due to space limitations, only the method of X is described below. X chart is shown in Figure 2, the vertical axis is the average of measurements of each verification, the abscissa is the verification time t, the baseline is the process mean, upper control limit (UCL) and lower control limit (LCL) are determined by the process parameters:

$$UCL = \bar{x} + k \cdot s_c \tag{6}$$

$$LCL = \bar{x} - k \cdot s_c \tag{7}$$

Where, k is the confidence factor, k=2 or 3, when k=2, the upper and lower limits is the warning limit, when k=3, it is the control limits.



**Fig. 2.** Average control chart

After accomplishing control charts, daily verification should be done in accordance with the defined time in the verification provisions. Data processing should be timely conduct after each verification, to calculate various statistics and draw them on the chart, to check whether the state is in the control limits. When the verification data always falls within the control limits, then the measurement process is controlled; when the data beyond the alert limit, it is necessary to analysis timely and take preventive measures; when the data beyond the control limit (as shown in Fig 2 on  $t_1$ ), and further verification to prove that the phenomenon is not accidental, then the measurement process is out of control.

Process parameters are usually revised in the following cases: 1) when the control limits is error; 2) when the verification standard is replaced; 3) after a period of time of verification, to correct the process parameters and update the baseline and upper and lower limit with available data. Amendments have two situations. When the initial parameter values and the large amounts of data from subsequent verification are basically same, recalculate the parameters according to all data of verification, the revised process parameters have a higher degree of confidence; when the initial data is inconsistent with verification data later, then the past history data void, re-calculate the new process parameters with the data obtained later. The verification interval should be further confirmed through inspection. The time interval of verification depends on the control situation of the measurement process, the frequency of verification should be higher generally at the beginning, the frequency can be reduced appropriately after the measurement process obtain a certain degree of confidence.

A program file should be established when the whole process is complete, the file content should cover with all requirements and procedures of verification.

### **Verification Method of Measuring Equipment**

Verification methods for measuring equipment include verification standards method, reserved samples comparison method, comparison method and so on[5]. We should base on the specific circumstances of the laboratory to select the appropriate verification methods.

Verification Standards Method. This method choose CRMs or measuring equipment calibrated by the higher measurement standards as the verification standard. Daily verification measure verification standard  $n$  times with measuring equipment which is to be verified, according to formula(1), calculate the average of measurements, measuring equipment calibration status is controlled if the results of each inspection meet formula(8).

$$\left| \overline{x_n} - A_s \right| \leq \sqrt{U_j^2 + U_s^2} \tag{8}$$

Where,  $U_j$ -the expanded uncertainty of  $x_n$ ;  $A_s$ -calibration value of verification standards (including reference material);  $U_s$ -expanded uncertainty of  $A_s$ ;  $U_s$  and  $U_j$  have the same coverage factor.

When standard uncertainty of CRMs or the verification standard is less than 1/4 of standard uncertainty of verification equipment introduced by the permissible error limit, determine equation simplifies to:

$$\left| \overline{x_n} - A_s \right| \leq \Delta \tag{9}$$

Where,  $\Delta$ -absolute value of the maximum permissible error of measuring device to be verified.

Reserved Samples Comparison Method. It is a verification method that detect again with reserved device under test (DUT), the DUT should have good stability, if necessary, with good resolution, low frequency of use. When check daily, calculate the average of measurements after  $n$  measurements of the reservation DUT, the calibration status of measurement standard did not change significantly if the results of each inspection meet the formula (10).

$$\left| \overline{x_n} - x_s \right| \leq \sqrt{2}U_s \tag{10}$$

Where,  $x_s$ -check reference value;  $U_s$ -the expanded uncertainty of  $x_s$ .

Comparison Method. Comparison method have two cases, one is that compare verification with two measurement equipment which have the same levels of accuracy. Typical test object with good stability should be used. Assume that the results of the compared verification were  $x_1$  and  $x_2$ ,  $U$  is the expanded uncertainty of measuring equipment to be verified. If the result satisfies (11), then the verifiable measurement standards state is effective.

$$\left| x_1 - x_2 \right| \leq \sqrt{2}U \tag{11}$$

The second is that measure the same stable object with three (sets) or over measurement equipments which have the same levels of accuracy.  $U_i$  is assumed as the expanded uncertainty of measurement equipment to be assessed.  $x_i$  is the measurement results measured by assessed measurement equipment,  $x_0$  is the average of measurements measured by more than one (sets) measuring equipment,  $x_i$  and  $x_0$  will be compared, if the formula (12) is satisfied, then the verifiable measurement standards state is effective.

$$|x_i - x_0| \leq \sqrt{\frac{n-1}{n}} U_i \quad (12)$$

### Verification Methods of Measurement Results

Repeat Testing with the Same Method. Repeat test use the same measurement methods on the same DUT to verify the measurement results in the original condition. It is useful when the results of previous tests are doubted, where require arbitration for a dispute.  $x_0$  is previously measured value;  $x_1$  is verification value;  $U_0$  and  $U_1$ , respectively, is the expanded uncertainty of  $x_0$  and  $x_1$ . If (13) is satisfied, the measurement results are consistent with measurement compatibility.

$$|x_0 - x_1| \leq \sqrt{U_0^2 + U_1^2} \quad (13)$$

Repeat Testing with Different Methods. We can also repeat the test using different methods on the same DUT, when the difference between two measurements satisfies (14), the measurement results are compatible.  $x_1$  is measurement from first method;  $x_2$  is measurement from second method;  $U_1$  and  $U_2$ , respectively, is the expanded uncertainty of  $x_1$  and  $x_2$ .

$$|x_1 - x_2| \leq \sqrt{U_1^2 + U_2^2} \quad (14)$$

Compare the Correlation Between Different Characteristics Measurement Results of the DUT. Some measure characteristics of the DUT have theoretical relevance, the credibility of the other characteristics measurement results can be inferred by one feature measurement, this correlation can be used to verify and analysis of the accuracy of measurement results.

## 4 Summary

Verification is an important mean to ensure the validity and reliability of measurement equipment, and the quality of test or calibration results which verification method would be choose is totally depending on the actual situation of different equipment test room, different testing equipment, but the the purpose and meaning of verification is the same. Through quantitative analysis of verification data, whether the measurement standards status of equipment is change can be grasped, to detect anomalies timely, to find the reasons, and remove faults, avoid the use of unqualified measurement standard device to affect the play of the overall performance of equipment. The methods of verification and standards of implementation summarized in this paper have a certain reference value for testing laboratories, calibration laboratories, and other measurement technology institutions' selection and application of verification methods.

## References

1. ISO/IEC17025-2005, General Requirements for Testing and Calibration Laboratory Ability (2005)
2. Ren, Z.-P.: During Check and Walter Control Chart. *Measuring Technology* (2005)
3. Gu, J.: How Carry out Period Checkin on Measurement Standards and Standard Equipments. *Measurement and Test Technology*, 69–70 (2008)
4. Zhu, J., Qin, K., Chen, S.-J.: Implementation of Intermediate Checks for Metrological Standard Equipment. *Instrument and Metrological Technology*, 25–28 (2006)
5. Ma, Y., Fu, B.: Discuss on Check Method for During Measurement Standard. *Railway Quality Control*, 32–33 (2008)
6. Williams, D., Marks, R., Davidson, A.: Comparison of on-Wafer Calibrations. In: 38th ARFTG Conference Digest, pp. 68–81 (1991)
7. Carullo, A.: Metrological Management of Large-Scale Measuring Systems. *IEEE Transactions on Instrumentation and Measurement*, 471–476 (2006)



# Fuzzy Control of Denitrifying Phosphorus Removal via Nitrite

Bin Ma<sup>1</sup>, Yongzhen Peng<sup>1,2</sup>, Shuying Wang<sup>2</sup>, Qing Yang<sup>2</sup>, Xiyao Li<sup>2</sup>,  
and Shujun Zhang<sup>3</sup>

<sup>1</sup> State Key Laboratory of Urban Water Resource and Environment,  
Harbin Institute of Technology, Harbin 150090, China

<sup>2</sup> Beijing Key Laboratory of Water Quality Science and Water Environment Recovery,  
Beijing University of Technology, Beijing 100124, China

<sup>3</sup> Beijing Drainage Group Co. Ltd., Beijing 100022, China

**Abstract.** To establish an fuzzy control system of denitrifying phosphorus removal via nitrite, a sequencing batch reactor (SBR) with on-line monitors of pH was applied to wastewater treatment. The results indicated that denitrifying phosphorus removal via nitrite could be achieved and inhibition of nitrite on phosphorus uptake could be avoided by step-feeding nitrite. During denitrification process pH increased with the decrease of nitrite, and the increase of pH was stopped after complete consumption of nitrite. So, an fuzzy control of denitrifying phosphorus removal via nitrite was established using pH as fuzzy control parameter. Further more, a novel two-sludge denitrifying phosphorus removal process was developed using above fuzzy control strategy.

**Keywords:** nitrite, denitrifying phosphorus removal, fuzzy control, two sludge, sequencing batch reactor (SBR).

## 1 Introduction

The nitrogen and phosphorus removal technologies aimed for eutrophication control have been the focus in the field of biological wastewater treatment. However, in the conventional biological nitrogen and phosphorus removal systems, the competition between denitrifiers and phosphorus accumulating organisms (PAOs) for substrate under anoxic conditions has been reported [1-3]. The denitrifying phosphorus removal was a feasible way to solve this competition. This technology could reduce the energy consumption (for aeration), the requirement of organic carbon source and the production of biomass[4,5].

Denitrification via nitrite could reduce the demand for carbon sources by 40%[6]. Therefore, if denitrifying phosphorus accumulating organisms (DPAOs) can use nitrite as an electron acceptor, more organic carbon source and energy (for aeration) can be saved [4]. It was reported, however, that the phosphorus uptake by DPAOs was seriously inhibited by the presence of a relative high level of nitrite (e.g. > 10mg NO<sub>2</sub><sup>-</sup>-N /L) [7,8]. In this experiment, the step-feeding of low level nitrite to a SBR was proposed to avoid this inhibition.

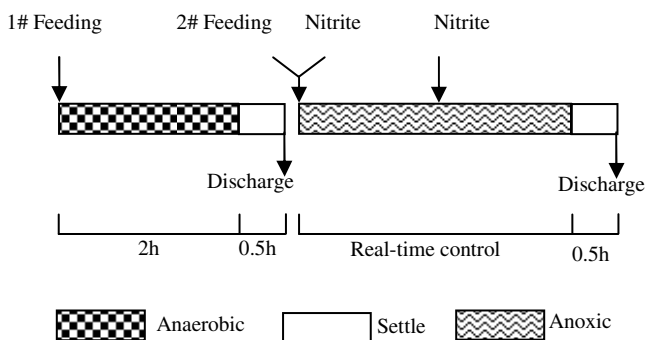
Fuzzy control has been successfully applied to wastewater treatment, especially in the biological phosphorus and nitrogen removal process. The activated sludge system is a complex dynamic system, so fuzzy control is a good choice for optimizing operation. The key is the use of simple and cheap on-line process measurements to infer the concentration of chemical variables, which are difficult or expensive to measure directly (nitrogen and phosphorus).

In this study, the relation of pH and nitrite concentration was investigated. Based on this information, an fuzzy control of denitrifying phosphorus removal via nitrite was established, and a novel two-sludge system with fuzzy control was developed.

## 2 Materials and Methods

### Reactor System

A laboratory-scale SBR made of plexiglass with a working volume of 12L was used in this study. An air-compressor was used for aeration. A mechanical stirrer was used to provide liquid mixing under anoxic phase. The pH, DO and temperature online sensor were installed for monitoring. The water temperature of the reactor was kept at  $24\pm 0.5^\circ\text{C}$  by a temperature controller. The seeding sludge was obtained from a SBR reactor working under alternating anaerobic-anoxic conditions using nitrate as electron acceptor.



**Fig. 1.** Operational strategy of SBR

Fig.1 shows the operation strategy of the SBR. Anaerobic and settle phases were fixed at 120 and 30 min durations, respectively. However, anoxic durations was not fixed but distributed by an real-time control strategy using pH as parameter to indicate the end of denitrification. Compared to the conventional SBR system, a settle phase was added between anaerobic and anoxic phase. 7.5L supernatant was removal at the end of this settle phase. Then, 7.5L synthetic wastewater without organic carbon source (2# feeding) and nitrite solution was instantaneously filled into the SBR reactor so as to avoid the effect of organic carbon source on denitrifying phosphorus removal. pH was adjusted at  $7.0\pm 0.1$  after each addition of nitrite by adding 1.0 M

HCl or 0.5 M NaOH. Mixed liquor suspended solids (MLSS) was 2218-2333mg/L in the SBR during the operation.

### Wastewater Composition

The 1# feeding was prepared by using tap water and the following composition: 384.4mg of  $\text{CH}_3\text{COONa}$  (300mg/L as COD basis), 49.4mg of  $\text{KH}_2\text{PO}_4$  (11mg/L as  $\text{PO}_4^{3-}\text{-P}$  basis), 112.5mg of  $\text{NH}_4\text{Cl}$  (30mg/L as  $\text{NH}_4^+\text{-N}$  basis), 14mg of  $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ , 90mg of  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ , 300mg of  $\text{NaHCO}_3$  and 0.3mL of nutrient solution per liter. The nutrient solution consisted of the following compounds per liter: 1.5g of  $\text{FeCl}_3\cdot 6\text{H}_2\text{O}$ , 0.15g of  $\text{H}_3\text{BO}_3$ , 0.03g of  $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ , 0.18g of KI, 0.12g of  $\text{MnCl}_2\cdot 4\text{H}_2\text{O}$ , 0.06g of  $\text{Na}_2\text{MoO}_4\cdot 2\text{H}_2\text{O}$ , 0.12g of  $\text{ZnSO}_4\cdot 7\text{H}_2\text{O}$ , 0.15g of  $\text{CoCl}_2\cdot 6\text{H}_2\text{O}$ , and 10g of EDTA [9].

The 2# feeding contained 224.5mg of  $\text{KH}_2\text{PO}_4$  (50mg/L as  $\text{PO}_4^{3-}\text{-P}$  basis), 112.5mg of  $\text{NH}_4\text{Cl}$  (30mg/L as  $\text{NH}_4^+\text{-N}$  basis), 14mg of  $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ , 90mg of  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ , 300mg of  $\text{NaHCO}_3$  and 0.3mL of nutrient solution per liter.

### Analytical Methods

The parameters measured including pH, temperature, DO, COD,  $\text{NO}_2^-\text{-N}$ ,  $\text{PO}_4^{3-}\text{-P}$  and MLSS. DO, pH and temperature were measured online using DO and pH meters (WTW340i, WTW Company). COD,  $\text{NO}_2^-\text{-N}$ ,  $\text{PO}_4^{3-}\text{-P}$  and MLSS were measured according to the standard methods given in APHA [10]. Samples were analyzed after filtration through 0.45  $\mu\text{m}$  filter paper.

## 3 Results and Discussion

The denitrifying phosphorus accumulating organisms (DPAOs) are able to store phosphate through sequential anaerobic-anoxic conditions. Organic carbon sources, particularly volatile fatty acids (VFAs), are taken up anaerobically and stored as poly- $\beta$ -hydroxyalkanoates (PHAs) through the release of phosphorus and degradation of glycogen. A higher amount of phosphorus is then taken up when nitrate or nitrite as an electron acceptor is supplied through PHAs oxidation, which is accompanied by biomass growth and the regeneration of glycogen. During the nitrite reduction, the alkalinity was produced, which resulted in the increase of pH. Based on this analysis, the relation between pH and nitrite concentration was investigated.

### Characteristics of Denitrifying Phosphorus Removal via Nitrite

The phosphorus release process in the anaerobic stage of denitrifying phosphorus removal via nitrite is the same as in the Fig.2 Typical SBR cycle conventional biological phosphorus removal process. So, this anaerobic phosphorus release process was not investigated in this study. The main goal is to show the relation between pH and nitrite concentration when nitrite is used as electron acceptor. The batch

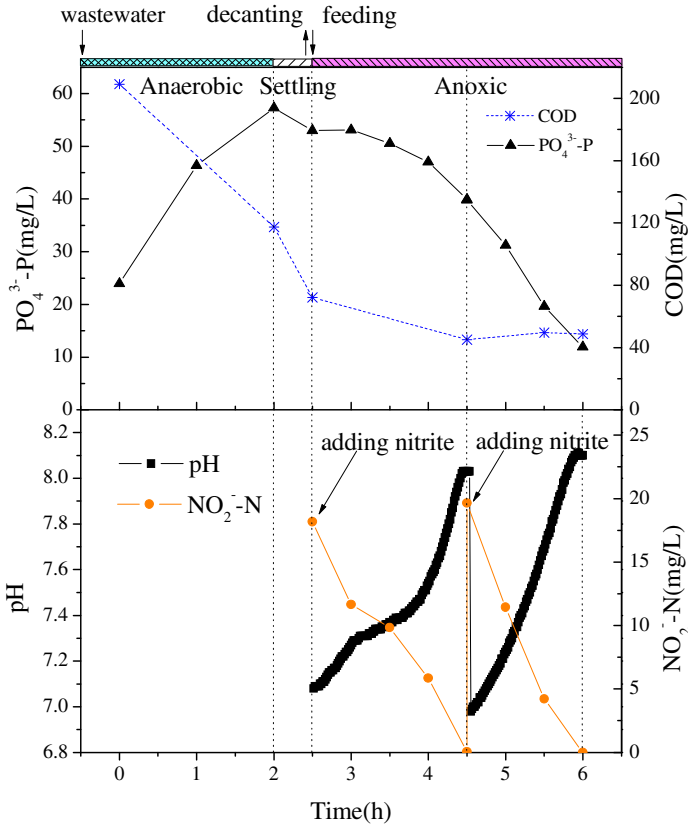


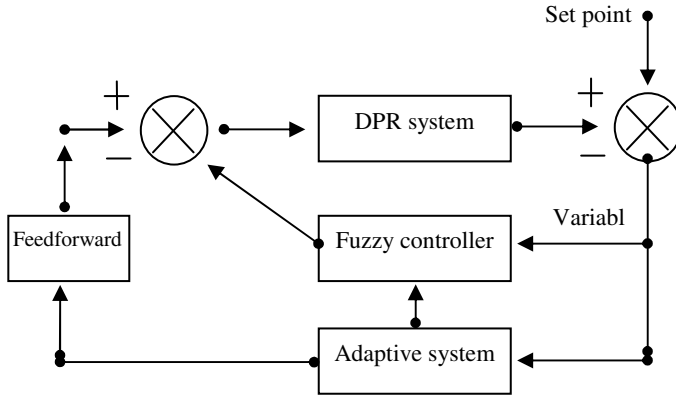
Fig. 2.

experiments were firstly conducted to determine the nitrite concentration which could not inhibit denitrification and phosphorus uptake. It was found that the inhibition of low level nitrite (below 20mg/L) was not obvious. So, the nitrite concentration in the SBR was controlled at about 20mg/L after adding nitrite. The characteristics of denitrifying phosphorus removal via nitrite was show in Fig. 2.

The nitrite concentration was 18.17mg/L after the first addition of nitrite. After 2 hours anoxic denitrification, the nitrite concentration decreased to 0.05mg/L. Meanwhile, anoxic phosphorus uptake happened with the phosphorus concentration decrease from 53.01mg/L to 39.94mg/L (Fig.2). It was also shown that pH increased to 8.03 and stopped after the depletion of nitrite. This result was due to the stop of producing alkalinity when the nitrite was completely consumed. After the second addition of nitrite, the variations of nitrite concentration, phosphorus concentration and pH were similar with that observed after the first addition. So there is an obvious correlation between pH variation and nitrite concentration variation. The process of denitrification can be inferred by on-line measuring pH of SBR.

### Establish an Fuzzy Control of Denitrifying Phosphorus Removal via Nitrite

To improve and optimize the stability of denitrifying phosphorus removal via nitrite, an fuzzy control strategy was established. This strategy include three parts (Fig.3): feed forward system, fuzzy controller and adaptive system.



**Fig. 3.** The strategy of fuzzy control of denitrifying phosphorus removal via nitrite (DPR. Denitrifying phosphorus removal)

The object of feedforward system was to provide approximate operational variables. The characteristics of raw wastewater has an important effect on the operation of wastewater treatment process, and this effect is nonlinear. So a lot of experiments and operational experience is indeed needed for establishing rules database. This feedforward system can give an approximate operation parameter which is benefit to obtain a stable performance.

Fuzzy controller was developed using two inputs, which were the pH error ( $E_{pH}$ ) between pH set-point and pH measured value in the SBR and the change of the pH error ( $CE_{pH}$ ). Out put variable is to add nitrite or stop mixing. Above results showed that pH was not changed after the complete consumption of nitrite, which meant that  $E_{pH}=0$  and  $CE_{pH}=0$ . However, Guo et al reported that pH began to decrease when the denitrification was stoped in the nitritaiton SBR[11]. So the nitrite also should be added into SBR when  $E_{pH}<0$  and  $CE_{pH}<0$ .

The adaptive system is used to calibrate feedforward system and fuzzy controller ,and to build a forecasting fault analysis system. Based on the results of this study and the experience of wastewater treatment, neural network and expert system were used to develop the forecasting fault analysis system, so as to reduce or avoid the fault occurrence.

### Developing an Novel Two-Sludge System to Achieve Denitrifying Phosphorus Removal via Nitrite

To achieve denitrifying phosphorus removal via nitrite, ammonium in raw wastewater should firstly oxidized to nitrite, where ammonium oxidizing bacteria (AOB) is

needed. After then, DPAOs is needed to conducted denitrification and phosphorus uptake. The requirements for AOB and DPAOs are different. So these two bacteria are seperated into two reactors, which benefit for optimizing nitritaiton and phosphorus removal, respectively. Based on this analysis, an novel two sludge procee is developed (Fig.4).

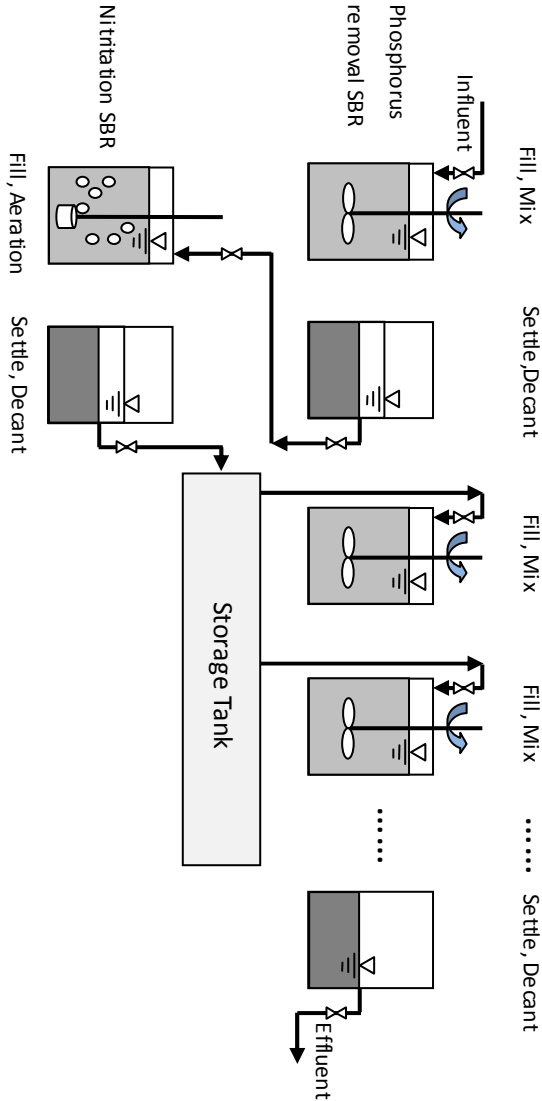


Fig. 4. The scheme of two sludge process achieving denitrifying phosphorus removal via nitrite

This two sludge system consists of two SBR reactors: phosphorus removal SBR and nitrification SBR. Wastewater is firstly filled into phosphorus removal SBR, where phosphorus is released to wastewater form activated sludge. Then supernate of phosphorus removal SBR is introduced into nitrification SBR, where ammonium is oxidized to nitrite by AOB. After nitrification supernate of nitrification SBR is decanted to the storage tank. Then wastewater is step feeded in phosphorus removal SBR using fuzzy control strategy. Denitrification and phosphorus uptake were happened, so nitrogen and phosphorus is removed from wastewater.

## 4 Conclusion

In denitrifying phosphorus removal process, pH increase with the decrease of nitrite in SBR. When nitrite was completely consumed, the increase of pH also stopped. pH can be used as on line measuring parameter to establish fuzzy control strategy. Meanwhile, an novel two sludge system was developed to remove nitrogen and phosphorus from wastewater through denitrifying phosphorus removal via nitrite.

**Acknowledgments.** This research was financially supported by State Key Laboratory of Urban Water Resource and Environment (HIT) (QAK201006) and Natural Science Foundation of China (50808004).

## References

1. Kuba, T., VanLoosdrecht, M.C.M., Brandse, F.A., Heijnen, J.J.: Occurrence of denitrifying phosphorus removing bacteria in modified UCT-type wastewater treatment plants. *Water Research* 31(4), 777–786 (1997)
2. Ahn, J., Daidou, T., Tsuneda, S., Hirata, A.: Characterization of denitrifying phosphate-accumulating organisms cultivated under different electron acceptor conditions using polymerase chain reaction-denaturing gradient gel electrophoresis assay. *Water Research* 36(2), 403–412 (2002)
3. Hu, Z.R., Wentzel, M.C., Ekama, G.A.: Anoxic growth of phosphate-accumulating organisms (PAOs) in biological nutrient removal activated sludge systems. *Water Research* 36, 4927–4937 (2002)
4. Zeng, R.J., Saunders, A.M., Yuan, Z.G., Blackall, L.L., Keller, J.: Identification and comparison of aerobic and denitrifying polyphosphate-accumulating organisms. *Biotechnology and Bioengineering* 83(2), 140–148 (2003)
5. Kuba, T., VanLoosdrecht, M.C.M., Heijnen, J.J.: Phosphorus and nitrogen removal with minimal cod requirement by integration of denitrifying dephosphatation and nitrification in a two-sludge system. *Water Research* 30(7), 1702–1710 (1996)
6. Peng, Y.Z., Zhu, G.B.: Biological nitrogen removal with nitrification and denitrification via nitrite pathway. *Applied Microbiology and Biotechnology* 73(1), 15–26 (2006)
7. Meinhold, J., Arnold, E., Isaacs, S.: Effect of nitrite on anoxic phosphate uptake in biological phosphorus removal activated sludge. *Water Research* 33(8), 1871–1883 (1999)

8. Lee, D.S., Jeon, C.O., Park, J.M.: Biological nitrogen removal with enhanced phosphate uptake in a sequencing batch reactor using single sludge system. *Water Research* 35(16), 3968–3976 (2001)
9. Smolders, G.J.F., Vandermeij, J., Vanloosdrecht, M.C.M., Heijnen, J.J.: Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process. *Biotechnology and Bioengineering* 44(7), 837–848 (1994)
10. APHA. *Standard Methods for the Examination of Water and Wastewater*, 19th edn. American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC, USA (1995)
11. Guo, J., Yang, Q., Peng, Y., Yang, A., Wang, S.: Biological nitrogen removal with real-time control using step-feed SBR technology. *Enzyme and Microbial Technology* 40(6), 1564–1569 (2007)



# Application and Challenge of It Service Management to University Informatization

Hong Lin and Yajuan Sun

Network and Information Center, North China Electric Power University, Beijing, China  
{l inh, syj}@ncepu.edu.cn

**Abstract.** This paper conducts a planning and designing of the IT service management strategy, its process, and the corresponding organizations on the basis of university informatization, and in the end builds a full-life-cycle service management system for university informatization.

**Keywords:** IT service management, ITIL, process.

## 1 Introduction

Statistics shows that the life cycle of an IT system is about 5-7 years, of which the system construction stage is about 1 year, while the system operation stage lasts over 4-6 years, which is the key stage in the life cycle of any IT system. Moreover, Gartner's survey finds that of all the IT system faults, those caused by the technology or products problems actually only accounted for 20%, management process failure problem accounted for 40%, and personnel error problem accounted for 40%, while as a matter of fact, personnel errors root from the absence of management. Thus, the IT service management system is very important for business and organization.

Some IT companies like IBM, HP, Microsoft has put forward corresponding ITIL model and solution, but these models' designing and developing are mainly based on the company's own products and service, which means the service targets at the enterprise users. Making profits is the ultimate goal of any enterprise, while University is a public welfare undertaking. This difference between university and enterprise determines the IT service management pattern designed by enterprises cannot be directly introduced into university. With regard to the university informatization, IT service management basically is a blank page, with no mature IT service management mode to follow, nor typical precedents of successful application to refer to. In the field of university informatization, the design and implementation of IT service management is a task which needs continuous exploration, research and innovation.

In the process of university's development, University Informatization is playing an increasingly important role in supporting its overall development. While the above problem becomes even more serious to the development of university informatization compared with the enterprises, for the neglect of IT service management has started to

restrict the development of university informationization and even university. In view of this, this article will conduct a study in IT service management as an important part of research in university informatization.

## **2 Nolan Model -- Developmental Phase of Informatization**

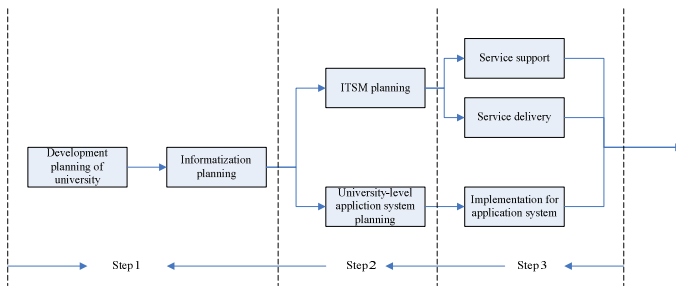
In the last 80's, Richard Nolan, professor of America Harvard University, after an in-depth study of the process of informatization, summarized a general rule of informatization development: for an industry, or for a country or region, informatization generally have to experience the developmental stages of initializing, expanding, controlling, integrating, data managing and growing mature, within which each stage is not separate from each other, but cannot be transcended. This is the famous Nolan model. [1]

A lot of revelation can be inspired from the analysis and introduction of the developmental phase theory of informatization, which are: the development of informatization basically follows the rule of growing from low level to high level, from part to whole, and from tactical pattern to strategic pattern. The construction of Informatization is not a once-for-all thing. It has different contents in different period, and its content is constantly changing, which require us carry out the information planning and construction with a view of development.

In the age of information technology, an important task of IT is to use IT technology, combined with the strategic goal of business system, constantly developing and perfecting new businesses, and improving the business value of service organization, so as to enhance its core competitiveness. Therefore, it is far less than enough only to focus on investment in the development and construction of system. More emphasis must be laid on IT service management, so as to ensure the all-round performance and value of service system.

## **3 IT Service Management Strategy for University**

The IT service management (ITSM) strategy is an important component of informatization strategy. Fundamentally speaking, when the IT service management planning is advanced by the University IT Department, it should serve the university's development planning. But as university development planning puts more emphasis on the macroscopic description of University 's long-term vision and mission, which belongs to strategic-level strategy that can not directly guide IT service management, thus three steps must be followed to incorporate the IT service management planning into the university development planning. And in this process, the IT service management planning and college application system planning should complement and support each other, as is shown in figure 1.



**Fig. 1.** The University IT service management strategy

Step 1: to draw up the informatization Development planning based on university's Development Planning.

Step 2: to synchronically complete the university application system (Strategic Information System) and IT service management planning.

Step 3: to synchronically implement the application systems and IT service management process.

According to the Nolan model, currently most colleges and universities are in the fourth stage of information development -- integration phase, focusing on the task of constructing the field application system (or digital campus) based on school top design, which will eventually provide teachers and students with university-oriented, unified one-stop integrated information service, whose basic realization consists of a whole process integration including hardware integration, data integration, service integration and service integration, and the gradual building and construction of an informatized atmosphere, informatized infrastructure, and informatized application environment meeting the requirement of a high-level research university.

The construction of field application system is a top-down design which involves the whole operations and process of school, so the system is an extremely large and complex one upon which all the university businesses depends. Therefore, the system has high requirements for being mature and stable. But practice has shown that, as a result of long-term neglect of IT service management in terms of investment, technology and management, IT departments in university are inexperienced in the operation and management of information system, resulting in the fact that the already-built system can not be operated into full play, which directly affect the application effect of informatization, and restricted the informationization and development of university.

## 4 Application of It Service Management in University Informatization

### Five Developmental Stages of IT Service Management

The key task of IT service management is to strengthen the ability to manage and control IT, and to build a perfect IT service management system. ITIL (IT Infrastructure

Library, IT Infrastructure Library) offers a theoretical framework based on best practice, with a theme consisting of six main modules, and a core of service management, as is shown in figure 2. International Organization for Standardization developed the international standard ISO / IEC20000 on the basis of ITIL. ITIL as a best practice, the ISO20000 as an international standard, they are both important basis for the management of IT services, what we lack now is how to successfully apply the standard and experience to practice and play a role.

According to the strategic vision of IT service management as well as the developmental need of its service object, IT service management can be divided into five developmental phases from low level to high level.

Stage1: “ruled by man”- A management model relying on the IT personnel, with IT personnel acting as firefighters, without process management tools.

Stage2: Passive- service station, incident management, problem management, change management, configuration management.

Stage3: Active- capacity management, availability management, release management.

Stage4: Service- A IT service oriented management model.

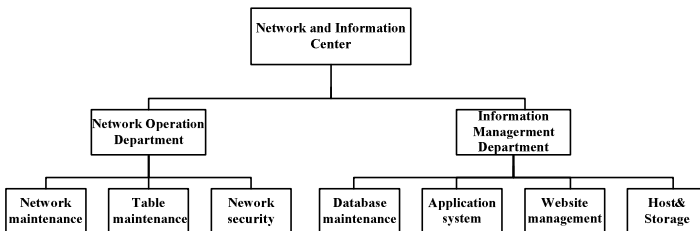
Stage5: value- as measured by business.

Of which the two stages of “passive service” and “active service” are process-driving stages. At present, University IT service management is in the primary stage of developing from "ruled by man" to process management.

**Construction of University IT Service Management System**

According to the basic framework of ITIL, IT service management can be divided into one management function and ten key processes. In the IT service management process, process drives IT operation for the task, but in the real work of university informatization, although the corresponding process has been established based on the ITIL framework, it often can not play its role, and even complicate the business, which reduces the IT operation efficiency. Analysis of the reason indicates that one of the important factors is that the organizational structure of IT department doesn't match its work processes.

Organizational structure of University IT department is generally divided into three layers in accordance with the business, as is shown in figure 2.



**Fig 2** structure of University Information Center

This organizational structure strictly divides the business ownership, and makes a clear division of labor and boundaries within the Department. But when facing the user, the structure forms a business barrier as it first requires the user to know the business division and post duty within the IT department, followed by the cross business problem, whose result may be the situation of mutually making excuses, thus the work efficiency is reduced.

The plan and design of IT service process should be built on a full understanding of the organizational structure of IT department. A comprehensive consideration of the process and business division must be given, as the construction of process needs the organizational structure to be adjusted, while the construction of process will also promote the adjustment of the organizational structure, which facilitates the service process system operates reasonably under the framework of organizational structure, and improves the management level of the organization.

According to the above analysis, an IT operation and maintenance system is designed combining IT service processes with the organizational structure of information center, which enables the University Information Center to be an integrated IT service center as a collection of information construction and management, whose three-layer structure is clear and reasonable. On the basis of this, according to the ITIL best practice, the construction of IT service system is under way.

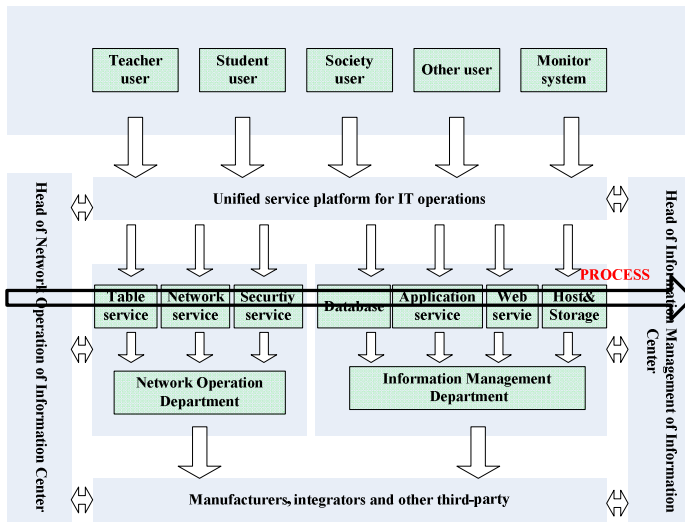


Fig 3. The University IT service society management system architecture

Referring to the ITIL best practice standards, combined with the University IT service management strategy and the organizational structure of information center, the IT operation and maintenance system is designed as a “three line” service system architecture. Horizontally, information center, the corresponding business departments and the IT service are matched, and businesses are incorporated together, breaking the tradition of business separation. In this way, an IT operation and

maintenance process system featuring Information Center is established. Longitudinally is the IT service architecture making up the three-line IT services, to provide support for IT operation and maintenance service.

The three-line operation and maintenance service system of university Information Center:

“Line 1”(service desk): Dealing with the service request and service consultation from the users; completing the distribution, supervision and feedback of service “event”; when “line 1” fails to provide support, transferring the problem to “line 2”, tracking the implementation and giving timely feedback to users.

“Line 2”(operation, management and technical personnel in information center): Converting problems from “line 1” or internal requests into events, accidents, problems, known errors, changes, or release, and accidents, problems, known errors, changes, or release, and implementing them through process; when “line 2” fails to offer a solution, transferring the problem to “line 3” for support.

“Line 3”(IT service provider):Providing corresponding technical support, based on service level agreement or contract; coordinating and processing when necessary.

### Full-Life-Cycle IT Service Management Process Planning

Though the IT service management has 11 processes, each process cannot function separately, but need to cooperate with each other and work together, forming a full life cycle operation. IT services should be like goods on the assembly lines, bringing high efficiency and high quality IT service through IT service management process. The drawback of users calling the Engineer directly can be eliminated by the combination of assembly-line-type service and reasonable organization structure of IT service, through a unified service platform, and with a new management system. Different faults and service request, after being identified by the service station and distributed into the event management, gradually transform into events, accidents, problems, known errors, changes, and release. Under this operation, the life cycle of a service starting from incident occurrence to ending as incident elimination is completed.

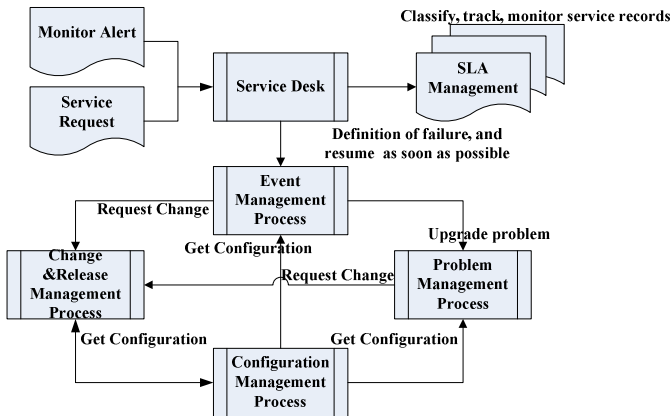


Fig 4. The full life cycle of the University IT service management process

## 5 Summary

Based on many years of working experience in IT operation and maintenance, the authors are committed to building a practicable IT management system which can be applied to the practice of university informatization. After the determination of IT service management strategy and the combination of the IT service management processes with the organization structure of information center, the final design of the full life cycle of IT service management processes and system structure is advanced. Through the above work, on the basis of the rational design of IT service management process, we constantly adapt to the needs of business development, thus realize the promotion and protection for the university informatization construction.

The IT service management is a dynamic process no matter in terms of process designing, or process application, so there is no unchangeable and permanent system. Therefore, it needs to be continuously adjusted, revised based on the development of business system, in order to achieve the best results.

## Reference

1. Gao, F.: The information resource planning, p. 14. Tsinghua University press (May 2005)

# A Novel Method for Iris Recognition Based on Zero-Crossing Detection

Li Yu and Xue Zhou

School of Automation, University of Electronic Science and Technology of China  
{lyu, zhouxue}@uestc.edu.cn

**Abstract.** As a kind of biometrics, iris recognition has been developed quickly in recent years. This paper studies a kind of iris identification algorithms which uses the longest-chord method for pupil localization and Hough transform for outer iris boundary location. Then the zero-crossing detection algorithm is used for iris feature extraction, and improved hamming distance is calculated for feature matching between templates of iris. Then, the influence of various parameters on the performance of the iris recognition program is studied deeply.

**Keywords:** Iris recognition, zero-crossing detection, hamming distance, feature extraction.

## 1 Introduction

Due to the higher requirements for security and personal privacy protection, biometrics has become a very good approach for setting security issues. Many kinds of biometrics have been applied in a wide range of modern society. Among all kinds of biometrics, iris recognition is one of the most accurate approaches to other biometrics.

Based on the feature extraction area of iris, the approaches for iris recognition can be divided into two categories. One kind of approach extracts the iris feature in the whole iris image, and the other one extracts the feature in the local area of iris image. Many researchers have proved that the later kind of method has better performance than that of the former kind of method. And many principal iris recognition methods use the local feature for iris recognition. Such as Daugman's method, it encodes the iris texture pattern into a 256-byte iris code by using some 2-dimensional Gabor filters, and taking the Hamming distance to match the iris code [1, 2]. Wildes matches images using Laplacian pyramid multi-resolution algorithms and a Fisher classifier [3]. Boles et al, extract iris features using a one-dimensional wavelet transform, but this method has been tested only on a small database [4]. Ma et al. construct a bank of spatial filters to characterize the key local variations of iris [5]. Later, they proposed an improved method based on characterizing key local variations with a particular class of wavelets, recording a position sequence of local sharp variation points in these signals as features [6]. Yuan et al. propose a method based on phase consistency and character point matching [7-9]. All these methods can obtain good performance



for iris recognition, but they all have high requirements for iris image. If the capture image is occupied too much by eyelid, or disturbed by too much noise, the performance of iris recognition will drop quickly.

To deal with these problems, Tian proposed a method based on zero-crossing detector to overcome these influences [10]. This paper is intended to improve this method on iris location, iris feature matching, and the influences of parameters used in the method are also studied deeply.

This paper is organized as follows. Section 2 discusses the iris preprocessing for iris location. Section 3 and 4 describe iris feature extraction and matching, respectively. Section 5 reports our experimental results. Section 6 offers a brief conclusion.

## 2 Iris Preprocessing

Our image preprocessing includes two steps. First, the iris is localized to remove the irrelevant parts (such as eyelid, pupil etc.) in a captured iris image. Second, the iris is normalized into a rectangular block to reduce the deformation caused by variations of the pupil.

### Isolation of the Region of Interest in the Iris Image

A captured iris image is a 2-D array ( $M \times N$ ). The gray level of point  $(x, y)$  is described as  $I(x, y)$ . The region of interest is the part between the pupil boundary and the sclera boundary. We first use the longest-chord based method to determine the location of the pupil, then use a circular template to search the outer boundary of the iris. The detailed steps are as follows [11]:

1) Scan the binary iris image  $I'(x, y)$  in the horizontal direction to find the longest chord. Here  $I'(x, y)$  is the binary iris image of  $I(x, y)$ . Since the pupil is generally darker than its surroundings, the longest chord in  $I'(x, y)$  will be the diameter of the pupil. Thus the center coordinates of the pupil are:

$$\begin{aligned} x_p &= x + \frac{Dia(x, y)}{2} \\ y_p &= y \end{aligned} \quad (1)$$

where  $Dia$  is the longest chord, and  $(x, y)$  is the start point of  $Dia$ .

2) Because of the contraction of the pupil and the variations in the shooting angle, the pupil will not be a regular circle. In many cases, the longest chord may not be the true diameter. There also may be several chords of the same length as the longest chord. Therefore, to determine the iris center, we select  $k$  chords near the longest chord. These will have a length between  $Dia - \Delta d$  and  $Dia$ . We then take the mean value of the midpoint coordinates of  $k$  chords. That is:

$$\begin{aligned}
 x_p' &= \frac{\sum_{Dia'(x,y) \in [Dia-\Delta d, Dia]} (x + \frac{Dia'(x,y)}{2})}{k} \\
 y_p' &= \frac{\sum_{Dia'(x,y) \in [Dia-\Delta d, Dia]} y}{k}
 \end{aligned}
 \tag{2}$$

3) Calculate the exact parameter of the outer boundary of the iris using Hough transform in a certain region determined by the center of the pupil ( $x_p', y_p'$ ), and then locate the outer boundary of iris by Hough transform.

**Iris Normalization**

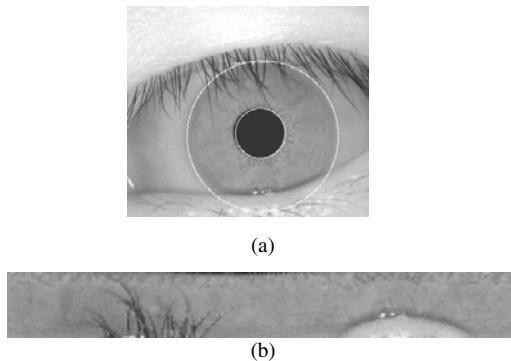
Once the pupil and sclera boundaries are extracted, the iris-ring, which is the part of the image between the two boundaries, has been located. As the captured iris image always varies in size, the detected iris is normalized into a rectangular block by the following mapping:

$$I(x(r, \phi), y(r, \phi)) \longrightarrow I(r, \phi)
 \tag{3}$$

where  $x(r, \phi)$  and  $y(r, \phi)$  are the linear combinations of points in the inner boundary ( $(x_{Inner}(\phi), y_{Inner}(\phi))$ ) and points in the outer boundary ( $(x_{Outer}(\phi), y_{Outer}(\phi))$ ):

$$\begin{cases}
 x(r, \phi) = (1-r) \cdot x_{Inner}(\phi) + r \cdot x_{Outer}(\phi) \\
 y(r, \phi) = (1-r) \cdot y_{Inner}(\phi) + r \cdot y_{Outer}(\phi)
 \end{cases},
 \quad r \in [0,1] \text{ and } \phi \in [0, \pi]
 \tag{4}$$

In the experiment, the preprocessed images are transformed into images of a size  $256 \times 32$ . Fig.1 shows the result of iris preprocessing.



**Fig. 1.** Preprocessing of iris image,(a) iris location (b) normalization or iris

### 3 Iris Feature Extraction

In order to reduce the influence of light, a gauss filter can be used to convolve the normalized iris.

$$I_{gauss}(x, y) = I(x, y) * G(x, y)$$

$$G(x, y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{4}$$

where  $\sigma$  is the size of gauss filter.

After filtering the image, a zero-crossing detect operator is used to convolve the filter image, then the binary code is obtained.

$$Detector_{zero-crossing} = [\underbrace{-1\dots-1}_n \underbrace{2\dots 2}_n \underbrace{-1\dots-1}_n] \tag{5}$$

Where n is the width of the zero-crossing detect operator, therefore, the zero-crossing detect operator is a 3n vector. The value of n is very small, and the sum of the detect operator is zero, so we can use this operator to extract local feature of iris.

$$Rslt(x, y) = I_{gauss}(x, y) * Detector_{zero\_crossing} \tag{6}$$

Here, the result of iris after convolution with the zero-crossing detect operator is  $Rslt(x,y)$ , the code of iris can be obtain by the following equation.

$$Code(x, y) = \begin{cases} 1, Rslt(x, y) \geq 0 \\ 0, Rslt(x, y) < 0 \end{cases} \tag{7}$$

Fig.2 shows the iris code from the same iris of Fig.1(b). From Fig.2, we can find that the iris code from the same iris has the similar code pattern.

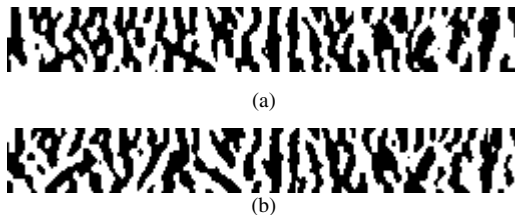


Fig. 2. Iris code from the same iris

Fig.3 shows the iris code from the different iris. It is obviously shows that the code pattern of iris from different iris is quite dissimilar.

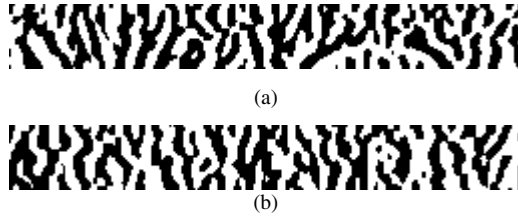


Fig. 3. Iris code from the different iris

#### 4 Feature Matching Based on Improved Hamming Distance

The eyelids, eyelashes and light often exist in the normalized iris image. If these points take part in iris feature matching, the result of iris recognition will become incorrect. Therefore, we should remove these points from iris image in feature matching process. An improved Hamming distance can be used for feature matching.

$$HD = \frac{\sum_{i,j} (TA \oplus TB) \cap (\overline{MA} \cap \overline{MB})}{\sum_{i,j} (\overline{MA} \cap \overline{MB})} \quad (8)$$

Here,  $TA$  and  $TB$  are two templates of different iris image,  $MA$  and  $MB$  are the disturbing masks of  $TA$  and  $TB$ .

Since the iris image is normalized into a rectangular block, the rotation transform in original images will change to a horizontal shift in the rectangular image. To avoid the influence of rotation transform, we can match an iris with several circularly shift iris image to get the minimum matching score as the final matching score.

$$HD_k = \frac{\sum_{i,j} (TA \oplus TB_k) \cap (\overline{MA} \cap \overline{MB}_k)}{\sum_{i,j} (\overline{MA} \cap \overline{MB}_k)} \quad (9)$$

$$HD = \min HD_k \quad (10)$$

Through comparing the HD with the threshold  $t$ , we can know that if  $HD < t$ , the two irises are come from the same iris. While  $HD > t$ , the two irises are come from the different one.

We use decidability  $d$  to evaluate the performance of our method.

$$d = \frac{|\mu_s - \mu_d|}{\sqrt{\frac{\sigma_s^2 + \sigma_d^2}{2}}} \quad (11)$$

Here  $\mu_s$  and  $\mu_d$  are the similarity mean values of the same class and different class, respectively. While  $\sigma_s$  and  $\sigma_d$  are the standard deviation values of the same class and different class, respectively. The bigger d is, the better decidability becomes.

### 5 Experimental Results

The proposed method for iris recognition is implemented in Matlab on a windows-XP PC with Pentium 2.4GHz processor and 2GB RAM. To evaluate the performance of the proposed method in verification mode, the public database CASIA 1.0 is used to test our method [12].

The proposed method has fast speed to locate the iris boundary. The average time for iris location is 0.57s, the maximum time is 0.86s, and the minimum time is 0.37s.

For the public database CASIA, the total number of same class comparisons is  $108 \times C_7^2 = 2268$  and that of different class comparisons is  $C_{108}^2 = 5778$ . Fig. 4 shows distributions of same class and different class matching distance HD in databases CASIA.

From the results shown in Fig. 4, we can find that the distance between the same class and the different class distribution is large, and the portion that overlaps between the same class and the different class is very small. This proves that the proposed features are highly discriminating.

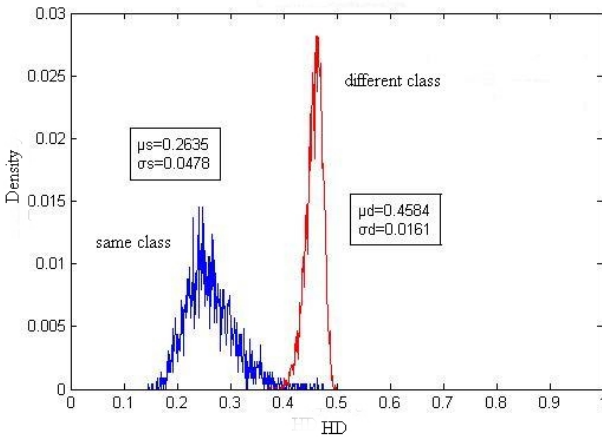


Fig. 4. Distributions of same class and different class distance

Fig.5 shows the curve of FRR (Fault Reject Rate) and FAR (Fault Accept Rate) with different threshold t. If the value of t becomes bigger, the FRR also becomes

bigger, while FAR becomes smaller. And if the value of  $t$  becomes smaller, the FRR also becomes smaller, while FAR becomes bigger. There is a tradeoff between FAR and FRR. In our test, the Equal Error Rate (ERR) equals to 1.06%.

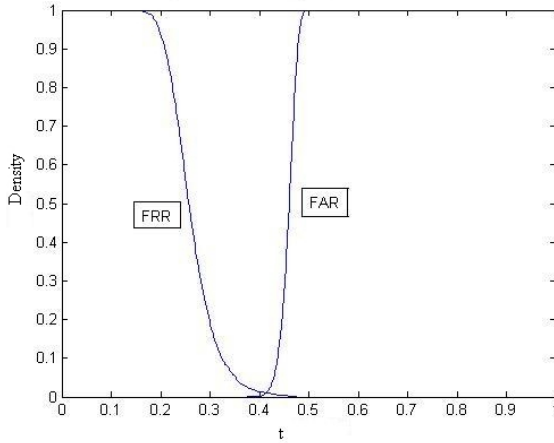


Fig. 5. FAR and FRR with threshold  $t$

Tab.1 shows the recognition results using different threshold  $t$ . When  $t=0.40$ , the proposed method can obtain the highest recognition rate.

Table 1. Recognition results using different threshold  $t$

$t$	FAR/%	FRR/%	$t$	FAR/%	FRR/%
0.35	0.0000	5.5996	0.43	5.1056	0.6614
0.36	0.0000	3.7037	0.44	12.3745	0.5291
0.37	0.0000	3.0423	0.45	25.1125	0.3968
0.38	0.0173	2.3369	0.46	46.6078	0.2205
0.39	0.0519	1.8519	0.47	72.7068	0.0882
<b>0.40</b>	<b>0.1038</b>	<b>1.3668</b>	0.48	92.3330	0.0000
0.41	0.5538	1.1905	0.49	99.3250	0.0000
0.42	1.8865	1.0141	0.50	99.9827	0.0000

Fig.6 shows the decidability with the width  $n$  of the zero-crossing detect operator. When  $n=2$ , the decidability of the proposed method becomes the largest.

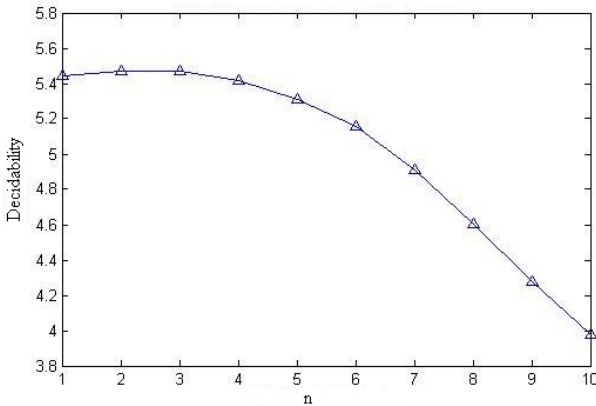


Fig. 6. Decidability with different number of shift n

We also test the influence of different size  $\sigma$  of gauss filter on the iris recognition, which is shown in Fig.7. When  $\sigma > 6$ , the decidability becomes stable. If the value of  $\sigma$  is too big, the filtering time of iris will be increased. Therefore, we make  $\sigma = 6$  in our tests.

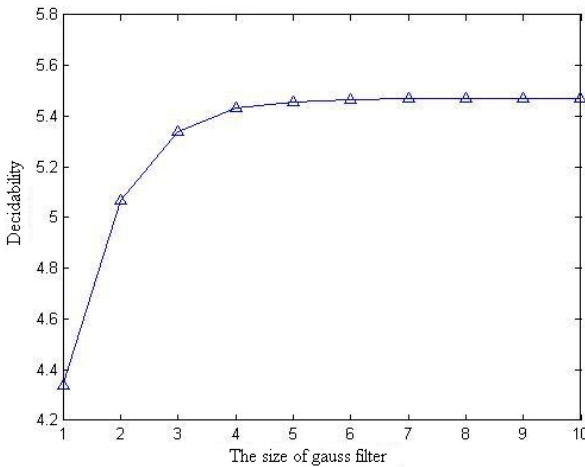


Fig. 7. Decidability with the size of gauss filter

## 6 Conclusion

This paper proposes a novel feature extraction method for iris recognition based on zero-crossing detection. The proposed method has been tested and evaluated on CASIA 1.0 databases. When the threshold value  $t$  equals to 0.4, the proposed method

can achieve the highest recognition rate 98.53%. And the selections of the parameters  $n$  and  $\sigma$  have also been discussed in detail to ensure the proposed method maintains a good compromise between accuracy and speed. In conclusion, the proposed method can distinguish an iris effectively and is a promising algorithm for establishing a real personal authentication system using iris biometrics.

**Acknowledgments.** The authors are very grateful for the support provided by the National Natural Science Foundation of China (60905015) and the Fundamental Research Funds for the Central Universities (ZYGX2009J068). Parts of the experimental data were supplied by the Institute of Automation, Chinese Academy of Sciences (CASIA) [12].

Yu Li- Associate Professor. The research interests are pattern recognition, image analysis, and biometrics, etc.

## References

1. Daugman, J.: High Confidence Visual Recognition of persons by a Test of Statistical Independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(11), 1148–1160 (1993)
2. Daugman, J.: How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology* 14(1), 21–30 (2004)
3. Wildes, R.P.: Iris recognition: An emerging biometric technology. *Proceedings of the IEEE* 85, 1348–1363 (1997)
4. Boles, W.W., Boashash, B.: A human identification technique using images of the iris and wavelet transform. *IEEE Trans. on Signal Processing* 46(4), 1185–1188 (1998)
5. Ma, L., Tan, T.-N., Wang, Y.-H., Zhand, D.-X.: Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(12), 1519–1533 (2003)
6. Ma, L., Tan, T.-N., Zhang, D.-X.: Efficient Iris Recognition by Characterizing Key local Variations. *IEEE Trans. Image Processing* 13(6), 739–749 (2004)
7. Yuan, W.-Q., Xu, L., Lin, Z.-H.: Iris recognition method based on phase congruency. *Chinese Journal of Scientific Instrument* 29(11), 2316–2319 (2008)
8. Yuan, W.-Q., Xu, L., Lin, Z.-H.: Adaptive iris recognition method based on texture feature point matching. *Chinese Journal of Scientific Instrument* 30(2), 233–236 (2009)
9. Yuan, W.-Q., Zhao, Y.-M., Zhang, Z.-J.: Iris recognition algorithm based on texture distribution feature. *Chinese Journal of Scientific Instrument* 31(2), 365–370 (2010)
10. Tian, Q.-C., Pang, Q., Cheng, Y.-M., Zhang, H.-C.: Iris Feature Extracting Algorithm Based on Zero-Crossing Detection. *Journal of Electronics & Information Technology* 08, 1452–1457 (2006)
11. Yang, W., Yu, L., Wang, K.-Q., Lu, G.-M.: A Fast Iris Location Algorithm. *Journal of Computer Engineering and Application* 40(10), 82–84 (2004)
12. CASIA Iris data version.1 [EB/OL] (October 2003), <http://www.sinobiometrics.com>



# Fast Iris Location Based on Window Mapping Method

Li Yu and Xue Zhou

School of Automation, University of Electronic Science and Technology of China  
{lyu, zhouxue}@uestc.edu.cn

**Abstract.** Iris recognition as one kinds of biometrics has become a hot research topic in recent years. As the first crucial step of iris recognition, the accuracy of iris location is very important for iris recognition. A novel iris localization method based on window mapping is presented here for iris segmentation: First, it searches the pupil center by window mapping, and then locates the pupil boundary. Finally, a fast and precise refinement of the outer boundary of iris is carried out based on an improved Hough transform method. Compared with conventional methods, this novel method shows faster and robust properties in locating the iris boundaries.

**Keywords:** Iris location, Hough transform, pupil center, iris recognition.

## 1 Introduction

With the high speed development of information technology, higher requirements for security and personal privacy protection are put forward in modern society. Biometrics is expected to become a very good approach for setting security issues. Among all kinds of biometrics, iris recognition has regarded as a competitive approach to other classical biometrics such as the fingerprint recognition, retina recognition, etc. However, the existing iris recognition algorithms have problems of slow operation and poor robustness. In order to improve the performance of iris recognition, many researchers put their efforts on iris location methods to improve the accuracy and speed of iris recognition.

Currently, the iris location methods mainly include: the integral-differential operator elaborated by Daugman [1-4], the Hough transform based operator elaborated by Wildes [5,6], least square method proposed by Wang [7] and the algorithm based on the gray-scale distribution feature [8,9], these methods all achieve good location effect. The analyses of these methods are all based on the circular approximated aspect of iris borders. Many researchers improved first two techniques to locate iris and to extract it from eye image [10-14]. Following a survey of these algorithms, we can conclude that the iris detection refers to accurately search respective radius of the two circles, as well as their respective centers are not concentric. Although many approaches can works with good results mentioned before, they are still failure to locate the iris edge when the iris is occupied by eyelids and eyelashes often having circular arc aspects and a random space distribution. Moreover, the reflection points particularly found inside pupil, often present obstacles. A fast and efficient determination of the pupil center and radius could

improve the performance to these approaches. It is the motivation of this work. We propose in the following a pupil localization strategy based on window mapping followed by an improved Hough transform to locate the outer boundary of iris. This method can fast localize the pupil center, which can speed up the iris boundary location. The experiments proved this method can greatly improve the speed and accuracy of iris location.

The rest paper is organized as follows: Section II will describe the details about the location of the pupil center and boundary with window mapping. Section III will introduce the details about the outer boundary location of iris based on the improved Hough transform. Section IV will evaluate the performance of this method. The conclusion of this paper is presented in Section V.

## 2 Location of Pupil Center and Boundary

The transcendent knowledge is used for pupil center searching. From an eye image, as show in Fig.1, we can find that the gray level of pupil is the lowest than other part of eye. To separate the pupil from the eye image, the original image is transformed to a binary image by using threshold segmentation. The analysis of the original image gray level histogram is displayed as Fig.2. The first peak of the histogram is a gathering of the pupil's gray value, while the second peak means the iris. So the threshold value should be a number just bigger than the value of the first peak.

A captured iris image is a 2-D array ( $M \times N$ ). The gray level of point  $(x, y)$  is described as  $I(x, y)$ . Then the binary image  $g(x, y)$  is obtained by the threshold T:

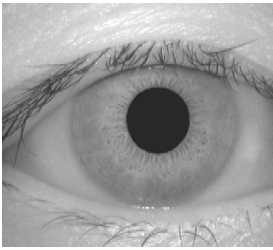


Fig. 1. Eye image

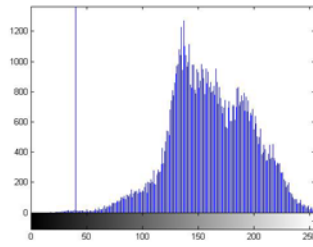


Fig. 2. Gray level histogram of eye image

$$g(x, y) = \begin{cases} 1, & I(x, y) < T \\ 0, & I(x, y) > T \end{cases} \quad (1)$$

Therefore, a square window mask can be used to find the pupil center. Suppose the mask is  $\text{Mask}(x, y)$ , the value of pixel in the mask is 1. Based on the experiences, we often use the average pupil diameter as the size of window mask.

$$Mask(x, y) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{size \times size} \quad (2)$$

Then, the binary image is convolved by the mask, the maximum value represents the possible location of pupil, then the center of the square mark can be regarded as the course pupil center. The window mapping for pupil center is showed in Fig.3.

$$(x_p^*, y_p^*) = (x, y) | \max(\bigcup_{x,y} Mask(x, y) * g(x, y)) \quad (3)$$

To calculate the radius of the pupil, the pupil boundary must be gained. In this paper, the pupil edge detection is realized by canny detector. Fig.4 shows the result of pupil boundary searching by canny operator.

Since the pupil center is coarsely located by window mapping, we can make a small area for searching the pupil boundary. Suppose the scope of pupil radius  $r_{pupil}$  is  $[r_{p \min}, r_{p \max}]$ , then we can search the edge of pupil in the area  $[x_p^* - r_{p \max}, x_p^* + r_{p \max}]$  and  $[y_p^* - r_{p \max}, y_p^* + r_{p \max}]$ .

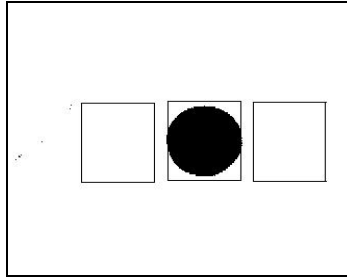


Fig. 3. Window mapping for pupil center location

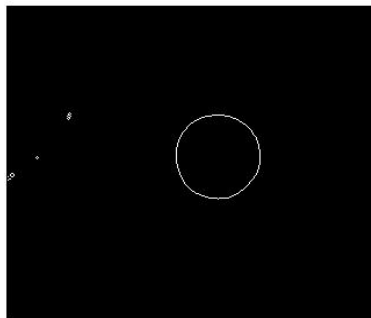


Fig. 4. Canny edge detection

In this stage, the edge points of this map are used for Hough transform to locate the edge of pupil. For a point whole picked out from the edge map  $(x_j, y_j)$ ,  $j = 1, \dots, n$ , the Hough transform is defined as follows:

$$H(x_p, y_p, r_{pupil}) = \sum_{j=1}^n h(x_j, y_j, x_p, y_p, r_{pupil}) \tag{4}$$

where  $(x_p, y_p)$  is the coordinate of pupil center.

$$h(x_j, y_j, x_p, y_p, r_{pupil}) = \begin{cases} 1, & g(x_j, y_j, x_p, y_p, r_{pupil}) = 0 \\ 0, & else \end{cases} \tag{5}$$

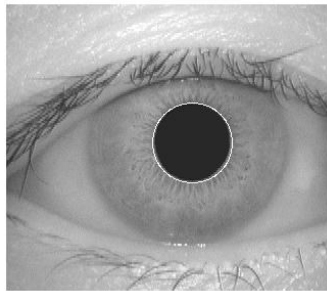
With

$$g(x_j, y_j, x_p, y_p, r_{pupil}) = (x_j - x_p)^2 + (y_j - y_p)^2 - r_{pupil}^2 \tag{6}$$

Based on the location of coarse pupil center, the coordinate of pupil center  $x_p$  and  $y_p$  can be reduced the reaching range. Therefore, the scope of  $x_p$  is in  $[x_p^* - \epsilon_x, x_p^* + \epsilon_x]$ , and  $y_p$  is in  $[y_p^* - \epsilon_y, y_p^* + \epsilon_y]$ , where  $\epsilon_x, \epsilon_y$  are the maximal distance between coarse pupil center and actual pupil center in row and column direction.

The triplet of parameters that maximizes  $H(x_p, y_p, r_{pupil})$  corresponds to a circle gathering the most number of edge points. Therefore, the pupil center and the radius can be confirmed by parameters  $(x_p, y_p)$  and  $r_{pupil}$ . Fig.5 shows the result of pupil boundary.

Our image preprocessing includes two steps. First, the iris is localized to remove the irrelevant parts (such as eyelid, pupil etc.) in a captured iris image. Second, the iris is normalized into a rectangular block to reduce the deformation caused by variations of the pupil.



**Fig. 5.** Location of pupil boundary

### 3 Outer Boundary Location of Iris

Like processing as the location of pupil boundary, we still use Hough transform to locate the outer boundary of iris. Since the upper and lower eyelids will possibly occupy part of iris image, the scope of iris image for edge detection can be reduced in a small area. Suppose the scope of iris radius  $r_{iris}$  is in  $[r_{i\min}, r_{i\max}]$ , then the edge of outer boundary of iris is limited in the area  $[x_p - x_{i\max}, x_p + x_{i\max}]$  and  $[y_p - r_{pupil}, y_p + r_{pupil}]$ .

The following steps can be used to reduce the amount of calculation of Hough transform and improve the accuracy of location.

Step 1: Using canny edge detector to search the edge of iris boundary. Unlike the processing of pupil boundary detection, we only use the horizontal gradient template of canny operator to search edge of iris.

Step 2: Since the pupil and iris center are not concentric, we can search the coordinate of iris center  $(x_i, y_i)$  in the scope  $[x_p^* - \delta_x, x_p^* + \delta_x]$  and  $[y_p^* - \delta_y, y_p^* + \delta_y]$ , where  $\delta_x, \delta_y$  are the maximal distance between pupil center and iris center in row and column direction.

Step 3: Using Hough transform to locate the boundary of iris. The Hough transform is defined as follows:

$$H(x_p, y_p, r_{iris}) = \sum_{j=1}^n h(x_j, y_j, x_p, y_p, r_{iris}) \quad (7)$$

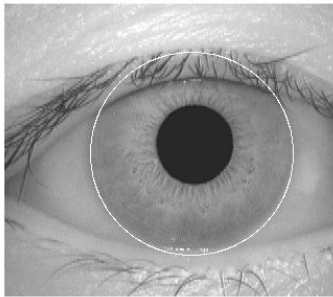
where

$$h(x_j, y_j, x_p, y_p, r_{iris}) = \begin{cases} 1, & g(x_j, y_j, x_p, y_p, r_{iris}) = 0 \\ 0, & \text{else} \end{cases} \quad (8)$$

With

$$g(x_j, y_j, x_p, y_p, r_{iris}) = (x_j - x_p)^2 + (y_j - y_p)^2 - r_{iris}^2 \quad (9)$$

Finally select the maximum value or equation (7), and its corresponding center point and radius is the iris center and the radius of outer boundary of iris. Fig. 6 is the location result of outer iris boundary.

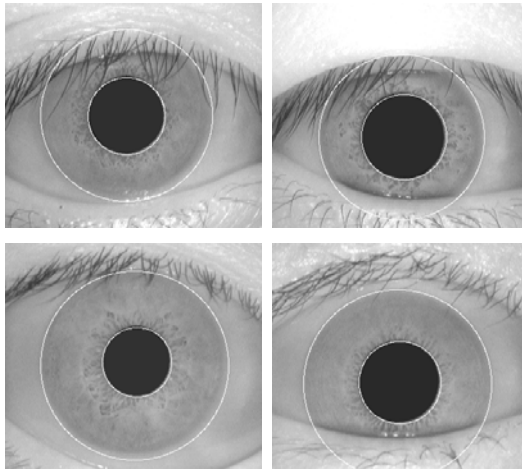


**Fig. 6.** Location of outer boundary of iris

## 4 Experimental Results

A new iris location method for iris segmentation is implemented in Matlab on a windows-XP PC with Pentium 2.4GHz processor and 2GB RAM. To evaluate the performances of the proposed method, the CASIA 1.0 [15] iris database is used.

There are totally 756 iris images in CASIA 1.0. Fig.7 shows some examples of the iris location results. From Fig.7 we can see that the proposed method can achieve high accuracy of iris location.



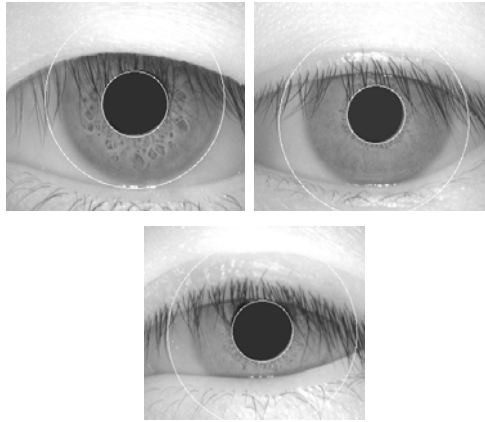
**Fig. 7.** Examples of iris location results of the proposed method

We also compare our method with other methods in Tab.1. The proposed method has the highest accuracy than other methods.

There are only three image are failure to locate the iris boundary. Fig.8 shows the failure examples.

**Table 1.** Comparison of performance of four methods

Algorithm	Hough Transform	Daugman's	Least square method	Proposed method
Accuracy	94.7%	98.85%	96.3%	99.6%

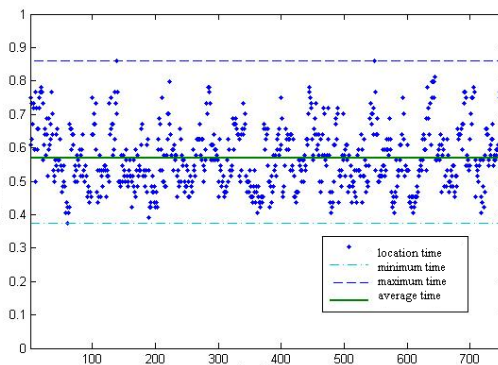


**Fig. 8.** Examples of failure iris location

The possible reasons of these examples are:

- (1) The contrast between iris and sclera is too small to find the edge of iris.
- (2) The occupy of the eyelid is too big, only a small iris area can be seen.
- (3) The boundary of eyelash will disturb the iris location result.
- (4) The boundary of iris has distorted too much.

The proposed method has fast speed to locate the iris boundary. The average time for iris location is 0.58s, the maximum time is 0.86s, and the minimum time is 0.37s. Fig.9 shows the time of iris location for CASIA 1.0. Because the proposed method uses window mapping to fast locate the pupil center and delete many useless points of iris image for boundary searching, the time of iris location can be greatly shortened.



**Fig. 9.** Time for iris location of CASIA 1.0

## 5 Conclusion

This paper proposed a fast iris location algorithm based on window mapping. This algorithm can fast locate the pupil center, which can help searching the edge of iris in small area. Experiments have proved it's valid in locating the boundary of iris. And the accuracy and speed of this method are also better than other traditional methods.

**Acknowledgments.** The authors are very grateful for the support provided by the National Natural Science Foundation of China (60905015) and the Fundamental Research Funds for the Central Universities (ZYGX2009J068). Parts of the experimental data were supplied by the Institute of Automation, Chinese Academy of Sciences (CASIA) [15].

Yu Li- Associate Professor. The research interests are pattern recognition, image analysis, and biometrics, etc.

## References

1. Daugman, J.: New Methods in Iris Recognition. *IEEE Trans. System, Man, and Cybernetics-Part B: Cybernetics* 37(5), 1167–1175 (2007)
2. Daugman, J.: How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology* 14(1), 21–30 (2004)
3. Daugman, J.: The Importance of being Random: Statistical Principles of Iris Recognition. *Pattern Recognition* 36(2), 279–291 (2003)
4. Daugman, J.: Statistical Richness of Visual Phase Information: Update on Recognizing Persons by Iris Patterns. *Int. J. Computer Vision* 45(1), 25–38 (2001)
5. Wildes, R.P., Asmuth, J., et al.: A Machine-vision System for Iris Recognition. *Machine Vision and Applications* 9, 1–8 (1996)
6. Wildes, R.P.: Iris recognition: An emerging biometric technology. *Proceedings of the IEEE* 85, 1348–1363 (1997)
7. Wang, Y.-H., Zhu, Y., Tan, T.-N.: Biometrics personal identification based on iris pattern. *Acta Automatica Sinica* 28(1), 1–10 (2002) (in Chinese)
8. Yuan, W.-Q., Xu, L., Lin, Z.-H.: Iris localization algorithm based on gray distribution features of eye images. *Journal of Optoelectronics Laser* 17(2), 226–230 (2006)
9. Yuan, W.-Q., Ma, J.-F., Di, W.-B.: A new method of iris location based on the active contour. *Computer Engineering and Application* 40(34), 104–107 (2003)
10. Vis, J.: Statistical richness of visual phase information: update on recognizing persons by iris patterns. *Int. Comput.* 45(1), 25–38 (2001)
11. Ma, L., Tan, T.-N., Wang, Y.-H., Zhand, D.-X.: Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(12), 1519–1533 (2003)
12. Ma, L., Tan, T.-N., Zhang, D.-X.: Efficient Iris Recognition by Characterizing Key local Variations. *IEEE Trans. Image Processing* 13(6), 739–749 (2004)
13. Wang, C., Li, F., et al.: Improved snake algorithm in the iris location application. *Application of Electronic Technique* 10, 31–33 (2006)
14. He, Z.-F., Tan, T.-N., Sun, Z.-N., Qiu, X.-C.: Towards Accurate and Fast Iris Segmentation for Iris Biometrics. *IEEE Trans. Pattern Analysis and Machine Intelligence* 31(9), 1670–1684 (2009)
15. CASIA Iris data version.1 [EB/OL] (October 2003)  
<http://www.sinobiometrics.com>



# The Principles of Disk Taint Analysis of Temu and How to Write a Plugin

Lanlan Qi<sup>1</sup>, Jiangtao Wen<sup>1</sup>, Bing Zhou<sup>1</sup>, Yu Chen<sup>1</sup>, and Zhiyong Wu<sup>2</sup>

<sup>1</sup> Dept. of Computer Science and Technology Tsinghua University, Beijing, China

<sup>2</sup> North Electronic Equipments Research Institute, Beijing, China

qll109@mails.tsinghua.edu.cn, jtwen@tsinghua.edu.cn,

zhoubing.tshcs@gmail.com, yuchen@tsinghua.edu.cn,

wuzhiyong@gmail.com

**Abstract.** This paper introduces dynamic taint analysis and the architecture of TEMU, and then, analyzes and summarizes the procedure and key functions of disk taint analysis, furthermore, this paper gives a plugin about disk taint and gives experiment results which verify the feasibility of using TEMU to implement disk taint analysis.

**Keywords:** disk taint, dynamic taint analysis, taint denotation, taint propagation, taint analysis.

## 1 Introduction

Taint analysis technique could be divided into static taint analysis technique and dynamic taint analysis technique. Taint analysis technique includes three steps: taint denotation, taint propagation and taint analysis [7]. Taint denotation normally marks the inputs of target programs as taint sources, taint propagation means the process when the taint sources flow and propagate in the program, and taint analysis implement analysis on this and abstract desired results.

Dynamic taint analysis has been successfully applied in detect cross script attack 3, SQL commands injection attack 56, buffer overflow attack 17.

TEMU89, proposed and developed by Heng Yin .etc, is a dynamic binary taint analysis platform based on emulator QEMU 10. Comparing other taint analysis techniques, TEMU has two advantages: (1) it provides a whole-system monitoring, and can trace both kernel-level and user-level codes without modifying running environment of the target program; (2) based on powerful emulator QEMU, TEMU can implement taint analysis on the operating systems running on embedded systems, and on applications on heterogeneous operation systems.

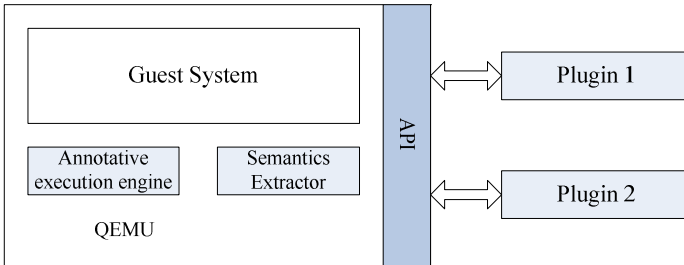
Taint analysis tools, like Valgrind 13, DynamoRIO 11, Pin 12 and PrivacyScope 14, are hard to implement denotation on physical disks, only by denoting memory after the data are read into memory, so they are hard to achieve full automation of implement taint analysis on disk. TEMU provides related interfaces to implement disk taint analysis, and this paper analyzes the mechanisms of disk taint analysis provided

by TEMU, explains key functions, gives a plugin about disk tainting and proposes a few existing questions.

## 2 The Principle of Disk Taint Analysis in TEMU

### 1 The Architecture of TEMU

TEMU is implemented based on QEMU which can implement dynamic binary taint analysis on target application, and two modules are added based on QEMU, one is annotative execution engine and the other is semantics extractor 8. TEMU provide plentiful API interfaces to users which can be used by users to write own plugins.



**Fig. 1.** The Architecture of TEMU

The author Heng Yin of TEMU provided three methods to implement taint propagation: shadow flag analysis, taint analysis and symbolic execution. The method of shadow flag analysis is according to execution environment of the context and semantics of current instructions we decide whether to mark memory addresses or registers.

The taint analysis in TEMU supports several taint sources like inputs from keyboard, network interfaces and hard disks. The basic taint propagation policy is: after a data source is tainted, every CPU instruction and DMA operation are monitored and are decided whether to propagate taint.

### 2 The Process of Disk Taint and Key Functions

TEMU v1.0 stores disk taint information by a whole array variable `disk_record_heads`, which can be seen as follows from `taintcheck.c`:

```

942 #define DISK_HTAB_SIZE (1024)
943 static LIST_HEAD(disk_record_list_head, disk_record)
944 disk_record_heads[DISK_HTAB_SIZE];

```

**Fig. 2.** The data structure used to store disk taint information

The way to computer index of the array given by the plugin is:

```
70 int64_t index = sect_num * 8 + sector_offset / 64;
```

**Fig. 3.** The way to calculate index

The size of every marked disk taint record is 64 bytes, like `uint64_t taint`; Normally, the size of a sector is 512 bytes, if we want to mark a sector, 8 records are needed.

TEMU marks the data in disk as taint sources by function `taintcheck_taint_disk()`, see in `taintcheck.c`:

```

947     int taintcheck_taint_disk(uint64_t index, uint64_t taint,
948                               int offset,
949                               int size, uint8_t * record, void *bs)
950     {
951         if(!TEMU_emulation_started) return 0;
952         #ifndef NO_PROPAGATE
953         struct disk_record_list_head *head =
954             &disk_record_heads[index & (DISK_HTAB_SIZE - 1)];
955         disk_record_t *drec, *new_drec;
956         .....
957         LIST_FOREACH(drec, head, entry) {
958             if (drec->index == index && drec->bs == bs) {
959                 found = 1;
960                 break;
961             }
962             if (drec->index > index)
963                 break;
964         }
965         .....
966         if (size2)
967             taintcheck_taint_disk(index + 1, taint2, 0, size2,
968                                   record + size * temu_plugin->taint_record_size,
969                                   bs);
970         #endif
971         return 0;
972     }

```

**Fig. 4.** The code of `taintcheck_taint_disk()`

Explanation: from line 1003 we know it is recursive function, and it's basic function is to check whether the taint source has been tainted (line 966-line 973), if so, find suitable shadow memory or corresponding index in `disk_record_heads` array to mark.

In disk taint analysis provided by TEMU, disk could be propagation mediums between memory and registers.

In TEMU, there are taint propagation between physical disks, physical memory and physical registers in guest OS. When data are read from physical disks and transferred into physical memory, emulator QEMU call function `bdrv_aio_read()` to transfer data from guest physical disk to IO\_buffer and call function `dma_buf_rw()` to transfer data from IO\_buffer to guest physical memory. And TEMU instruments function `taintcheck_chk_hread()` into `dma_buf_rw()` to implement taint propagation between guest physical disks and physical memory, see fig. 5.

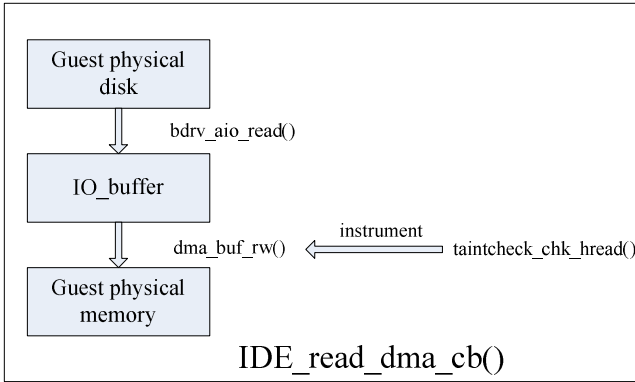


Fig. 5. Taint propagation from disk to memory

When data are read from physical memory into physical disks, `IDE_write_dma_cb()` call `dma_buf_rw()` to write data from guest physical memory into IO\_buffer, `IDE_write_dma_cb()` call function `bdrv_aio_write()` to write data from IO\_buffer into guest physical disk. While TEMU instruments function `taintcheck_chk_hwrite()` into function `dma_buf_rw()` to implement taint propagation from guest physical memory to guest physical disk, see Fig. 6.

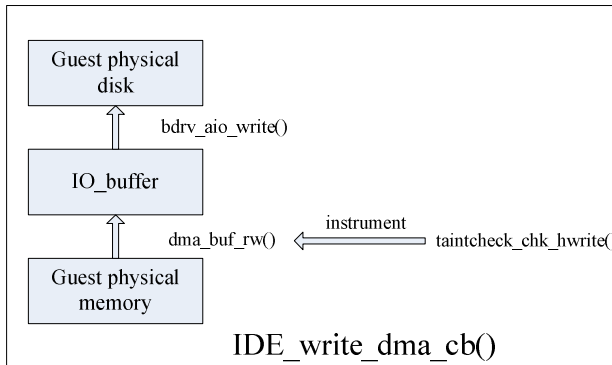


Fig. 6. Taint propagation from memory to disk

When CPU reads or writes disks, the taint propagations between guest physical disks and guest physical registers occurs. QEMU reads data and store them into guest disks by the function `ide_data_readw()` and the function `ide_data_readl()`, and writes data into guest physical disks by `ide_data_writew()` and `ide_data_writel()`. TEMU separately instruments `taintcheck_chk_hdin()` and `taintcheck_chk_hout()` into `ide_data_readw()` and `ide_data_readl()` to implement taint propagation between guest physical disks and guest physical registers, see Fig. 7.

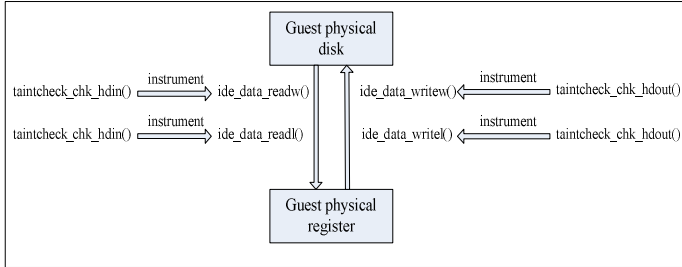


Fig. 7. Taint propagation between disk and register

### 3 Plugin and Experiment

#### 1 The Functions of the Plugin and Running Environment of Guest Host

Plugin information: the function of the plugin is to mark tainted data in disk and filter out the tainted codes in the target application.

Part of the source code can be seen in `./sample_plugin/main.c`:

```

69 void do_taint_file()
70 {
71 int sect_num = 13670828;//found from .img file
72 int sector_offset = 0;
73
74 //the algorithm of the index between addr
75 uint64_t index = sect_num * 8 + sector_offset / 64;
76 uint64_t taint = 1;//indicates the following byte is tainted
77
78 taint_record_t record_1;
79 bzero(&record_1, sizeof(record_1));
80
81 record_1.origin = 1;
82 record_1.offset = 0;
83 taintcheck_taint_disk((uint64_t)index, (uint64_t)taint, (int)0,
      (int)1, (uint8_t *) &record_1, (void*)TEMU_get_bs());
84 }

```

Fig. 8. The function of a plugin to mark taint source in the disk

Guest OS information: the guest OS is winxpsp2. And the target application is notepad.exe.

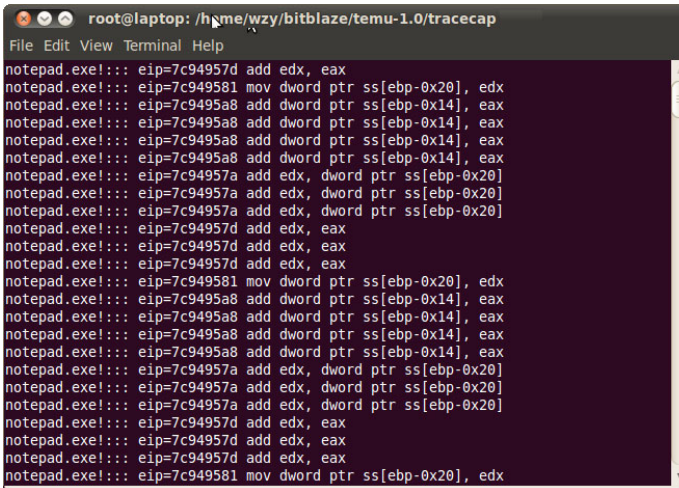
## 2 Run the Plugin

Execution of the plugin:

- 1) load\_plugin ../samples\_plugin/plugin.so
- 2) enable\_emulation
- 3) taint\_file
- 4) //open notepad.exe in guest OS
- 5) monitor\_proc notepad.exe
- 6) //use notepad.exe to open big file 3.txt

The tainted codes of notepad.exe are printed out, see Fig. 9

By debugging, the printed instructions are the tainted instructions of notepad.exe.



```

root@laptop: /home/wzy/bitblaze/temu-1.0/tracecap
File Edit View Terminal Help
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c949581 mov dword ptr ss[ebp-0x20], edx
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c949581 mov dword ptr ss[ebp-0x20], edx
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c9495a8 add dword ptr ss[ebp-0x14], eax
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957a add edx, dword ptr ss[ebp-0x20]
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c94957d add edx, eax
notepad.exe!::: eip=7c949581 mov dword ptr ss[ebp-0x20], edx

```

Fig. 9. Tainted instructions of notepad.exe

## 3 Notes about the Plugin

There are four notes about the plugin:

1) If the guest OS is windows, the physical address of the target file got by Winhex which installed in the guest OS is not accurate, we need to get it by other ways, see line 71 in main.c.

2) The first parameter and the second of taintcheck\_taint\_disk should be casted, or there will be error.

3) The last parameter of taintcheck\_taint\_disk need to be got by reconstructing function, and the plugin given by TEMU doesn't provide any interfaces.

4) The experiment target function should be big files, because in windows small files will be automatically put into OS caches, when target read this target file, it will read them directly from caches but not disk and thus the disk taint propagation failed. What's more, even put the small file in other partitions of the guest OS, it will fail too.

## 4 Summarization

This paper introduces dynamic taint analysis and the architecture of TEMU, explains the disk taint process provided by TEMU, and gives a plugin to verify the feasibility of using TEMU to implement disk taint analysis. Furthermore, this paper gives some notes when writes and uses the plugin.

**Acknowledgments.** This material is based upon work partially supported by the National Natural Science Foundation of China No 61070208.

## References

1. Kong, J., Zou, C.C., Zhou, H.: Improving Software Security via Runtime Instruction-level Taint Checking. In: ASID 2006: Proceedings of the 1st Workshop on Architectural and System Support for Improving Software Dependability, pp. 18–24 (2006)
2. Newsome, J., Song, D.: Dynamic Taint Analysis for Automatic Detection, Analysis, and Signature Generation of Exploits on Commodity Software. In: Proceedings of the Network and Distributed System Security Symposium, NDSS 2005 (2005)
3. Nguyen-Tuong, A., Guarnieri, S., Greene, D., Shirley, J., Evans, D.: Automatically Hardening Web Applications Using Precise Tainting. In: 20th IFTP International Information Security Conference (2005)
4. Qin, F., Wang, C., Li, Z., Kim, H.S., Zhou, Y., Wu, Y.: LIFT: A Low-Overhead Practical Information Flow Tracing System for Detecting Security Attacks. In: MICRO 2006: Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture, pp. 135–148 (2006)
5. Halfond, W., Orso, A., Manolios, P.: Using Positive Tainting and Syntax-aware Evaluation to Counter SQL Injection Attacks. In: SIGSOFT 2006/FSE-14: Proceedings of the 14th ACM SIGSOFT International Symposium on Foundations of Software Engineering, pp. 175–185. ACM Press, New York (2006)
6. Pietraszek, T., Berghe, C.V.: Defending Against Injection Attacks Through Context-Sensitive String Evaluation. In: Valdes, A., Zamboni, D. (eds.) RAID 2005. LNCS, vol. 3858, pp. 124–145. Springer, Heidelberg (2006)
7. Tang, H., Huang, S., Zhang, L.: Taint Propagation Analysis Algorithm for Vulnerability Exploits Detection. *Journal of Chinese Computer Systems* 31(11) (2011)
8. Yin, H., Song, D.: TEMU: Binary Code Analysis via Whole-System Layered Annotative Execution. In: Virtual Execution Environments 2010, Pittsburgh, PA (2010)
9. <http://bitblaze.cs.berkeley.edu/release/temu-1.0/howto.html> (2011)
10. <http://qemu.org/index.html> (2011)
11. Bruening, D., Garnett, T., Amarasinghe, S.: An infrastructure for adaptive dynamic optimization. In: International Symposium on Code Generation and Optimization, CGO 2003 (March 2003)
12. Luk, C.-K., Cohn, R., Muth, R., Patil, H., Klauser, A., Lowney, G., Wallace, S., Reddi, V.J., Hazelwood, K.: Pin: Building customized program analysis tools with dynamic instrumentation. In: Proc. of 2005 Programming Language Design and Implementation (PLDI) Conference (June 2005)
13. Nethercote, N., Seward, J.: Valgrind: a framework for heavyweight dynamic binary instrumentation. In: PLDI, pp. 89–100 (2007)
14. Zhu, Y., Jung, J., Song, D., Kohno, T., Wetherall, D.: Privacy Scope: A Precise Information Flow Tracking System For Finding Application Leaks. In: EECS (2009)

# Study for the Cooperation of Commercial Procedure Basing on Agent Technique

Shiyong Ding and Shuli Wang

Academy of Armored Forces Engineering, Fengtai, Beijing 100072, China  
zhangkei@sohu.com, wwz0528@sina.com

**Abstract.** The work procedure and role in the commercial procedure are analyzed, Agent technique is brought into the commercial procedure through the analysis and study for it, the model for the element is established, and the cooperative setup structure for the system is built, the analysis for cooperation process basing on Agent is also given.

**Keywords:** Agent, Commercial Procedure, Cooperation, Intelligence, Automation.

## 1 Introduction

The typical commercial procedure includes the disposition of supply chain, fixing travel in advance, meeting submission, estimation for production, demanding compensation and the bank business, etc., these processes involve some task of manual, semi-automatic or automatic work, need to address the different database, call external application, which present the complicated business logic, and the system becomes the complicated one. The modeling method basing on Agent supplies good thinking for reducing the system's complexity, and gives the system more flexibility, and cooperation.

## 2 Coordinated Commercial Procedure

The coordinated commercial procedure means not only the connection among the departments inside enterprise, but also the cooperative partners, supply merchants, retailers and even the terminal users, unifying the plan and data model for the dynamic union and coordination. All members of the commercial chain complete material purchasing, production designing, producing, sale & payment under the unified planning, they also scout for the latent market through their coordinated forecast. All members cooperate with one another around the order, including the information sharing and profession cooperation, so the mutual aid, fair and double benefit aspect is formed by them. With the quickly development of the information technique, the commercial process needs more and more technique for the automation and intelligence, and reduces the system's complexity. The Agent technique proves an important thought for these requirements.



### 3 Agent Techniques and Its Modeling Method

#### The Conception and Characteristics of Agent

There is no unified understanding yet for Agent in sphere of learning now, the differences come from the different research sphere. In sphere of information especially in computer and artificial intelligence, Agent means the computing entity that can run in some environment, respond the change of environment, and adopt some action for its designing target automatically.

Agent has some characteristics for itself besides the adaptation and cooperation that belong to the distributed artificial intelligent system:

A: The intelligent function. The Agent has the intelligence in a higher degree; it can be the model for describing the human intelligence and machine intelligence.

B: The opening system. Agent can not only perceps environment and makes reaction to it, it also can integrate other newly joining Agents into the system without the redesigning for the system, so it has the properties of opening and expansion.

C: Cooperative work. Each Agent can work independently; they also can work coordinately and cooperatively for solving the problem that the single Agent cannot solve concurrently. This can improve problem solving capability of the whole system.

D: Automatic action. Agent can control its action by its own, and plan its action according to the different targets or requirements from environment.

E: The target-oriented property. Agent can shows the action under the direction of some target and act automatically for realizing its initial target, besides the reaction for environment affairs.

#### Multi-Agent System

At present, the coordinated commercial modeling techniques for research are in varied forms, there are mainly enterprise modeling, network design method, approximation method and optimization simulation, etc. These methods all have their own good points, but they all exists the matching relation problem between the theory layer and application layer. Agent technique can not only supply the advantageous technique sustainment for the compact & concerted coordinated commerce, but also the reliable insurance for the new type of opening, reconfigurative and flexible commercial environment that is oriented to the distributed computation, all these functions come from the semantics mutual operation and action coordinative capability of Agent. And the network system-Multi-Agent System becomes the important technique for problem solving and intelligence improving of the commercial system. MAS has the good property of flexibility, which is suitable for solving the distributed problem of scattering data, knowledge and control. MAS stresses on the automatic decision and the coordinative capability of problem solving among Agents, these characteristics are suitable for the automation, distribution and

paralleling property of the coordinative commerce's actual running. Therefore, it becomes the important trend in the sphere of computer application and management science for introducing MAS into the coordinative commercial chain, which can simulate, optimize, realize and control the running of system.

## **4 The Basic Setup Structure for Commercial Process Basing on Agent**

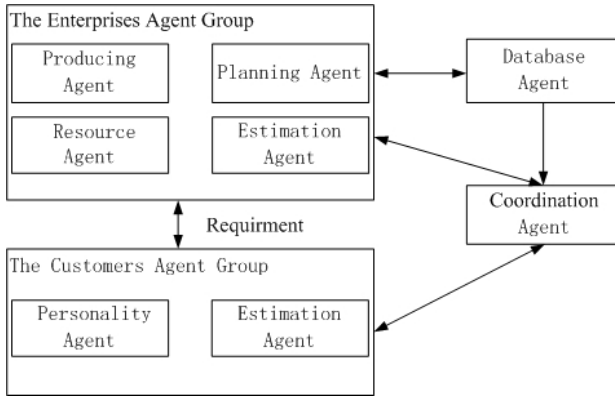
### **Analysis for Work Procedure and Role**

The commercial process is also a big complicated system as other systems of human society. In the basic procedure of this system, order in supply chain is the origination point. After the order is accepted, on the one hand corporation should get on to the enterprise for putting into producing plan and supervise the procedure & quality of the production, on the other hand, the corporation should arrange the commercial negotiations, business plan and other arrangement related to customer, and give attention to the capital and customer service. With the development of computer and the networks, the efficiency of commercial process is also improved quickly. The commercial process usually includes order of supply chain confirmation, product estimation, business arrangement ( meeting, travel, etc.) and other business related to customer service. Actually, there are only three matters in process: firstly, customers have their requirements, secondly, corporations focus on these requirements and give the necessity coordination, thirdly, supply the satisfying commodity service for customers.

According to the above-mentioned analysis about commercial process, there are three roles in it: the customer role, the sales agent role and the manufacturing enterprises role. The roles in the commercial process must be distinguished before using the Agent technique.

### **Setup Structure Basing on Agent of the System**

The definition of the roles in commercial system could strengthen the direction for the utilization of Agent modeling method. According to the system's running fact, the sales agent can be transformed into the coordination Agent because of coordination action between manufacturing enterprises and customers, other two groups could be transformed into enterprises Agent group and the customers Agent group. The enterprises Agent group includes planning Agent, resource Agent, producing Agent and estimation Agent, the customers Agent group includes the personality Agent and estimation Agent. The setup structure basing on Agent is indicated as in the figure 1. There is also a database Agent in the system for data & knowledge accumulation.



**Fig. 1.** Setup Structure Basing on Agent for Commercial Process

As fig.1 indicates that the external function frame and the initial data & information must be unified into one standard for realizing the cooperative function when using Agent modeling technique in commercial cooperative system. The function of system has been distributed in modularity, this can reduce the system's complexity, and improve system's cooperation capability & flexibility during the complicated commercial process. At the same time, the system basing on MAS can be easily expanded, that is to say, the functional entity that has the close function or same requirement could be integrated into the system, and so the system becomes an easy-expanded system with complete function and market variations adaptation.

## 5 Analysis for the Cooperation Procedure

The personality Agent in customers group puts out commodity requirement, and submits it to the cooperation Agent (the commercial corporation ), these generates the supply chain order. The cooperation Agent demonstrates and analyzes the requirement, and recovers the characteristics from it, then submits it to the enterprise Agent group. The group estimates the requirement characteristics, and after this, the suitable producing plan is formulated, the resource in the all group is called for putting into production, the Agents in the group cooperate one another effectively for completing the task. Finally, the production is submitted to the coordination Agent for estimation, the production is brought into commercial negotiation after the identification, then the series of negotiation for price is carried out until it is submitted to customers. The customers will also estimate the commodity (production), if they are satisfied with it, the commercial process is completed; otherwise, and there will be a circle until satisfaction

## 6 Conclusion

The automation and intelligence level will be greatly improved for the introduction of Agent technique. The further improvement needs the accumulation of knowledge and

experience. This also relies on the continuous regeneration of database knowledge, and the continuous perfection of the system's setup structure. With the overall opening and the deeply sharing of the data & information, it means much more commercial opportunity and service, and the cooperation becomes more and more complicated. The continuous mature and extensive application of the Agent technique can help the cooperation in commercial procedure reach to a new step.

## References

1. Guo, H.-X., Wu, J., Zhang, D.-J., Wang, C.-R.: Survey on Multi-Agent Technology. *Henan Science* 122(2), 242–246 (2004)
2. Ju, C.-H., Zhang, J.: Research on workflow model based on role-driven multi-agent. *Computer Engineering and Design* 27(21), 3979–3982 (2006)
3. Wu, Q., Yao, L., Liu, F.: The Study of e-Commerce Based on Multi-Agent Collaborating Technology. *Computer Engineering and Application*, 143–145 (January 2003)
4. Yao, L., Zhang, W., Gong, Y.: The Cooperated Work System Based on Multi-Agent. *Computer Engineering and Application*, 63–66 (December 1999)
5. Liu, G.-Q., Chen, X.-P., Fan, Y., Cai, Q.-S.: A Formal Model of Multi-Agent Collaborative Systems. *Chinese J. Computers* 24(5), 529–535 (2001)

## Author Biography

**DING Shiyong:** male, doctor, lecturer, 1973.11-, PengLai city, ShanDong province. E-mail: zhangkei@sohu.com

**WANG Shuli:** male, doctor, professor, 1962.06-, Weihai city, ShanDong province. Email: wwz0528@sina.com

# An Analysis and Assessment Method of WCDMA Network Traffic

Xinzhou Cheng and Jian Guan

Beijing Telecom Planning & Designing Institute Co., Ltd.  
Beijing, China  
{Chengxinzhou, guanjian}@btpdi.com.cn

**Abstract.** With the continuous expansion of network scale, how to balance the network quality and the construction cost has become the cruxes which operators focus on. Accurately evaluating wireless network traffic and rationally deploying network resources is the key to solve the problems above.

This paper analyzes the features of the WCDMA network traffic, and puts forward a new assessment method of WCDMA network traffic as a reference for the rational network resources allocation.

**Keywords:** WCDMA, traffic assessment, volatility, disequilibrium, resources allocation.

## 1 Background and Problems

With the continuous expansion of WCDMA service and technology innovation, the burst of services brings a greater impact onto the network. Traditional business forecasting methods lead to varying degrees of CE congestion due to inaccurate estimation. At the same time, there are also a number of base stations with lower utilization in existing networks, which implies the mismatch of resource allocation and total traffics.

The current wireless network capacity planning usually takes the deterministic analysis method based on business forecast analysis, network capacity is designed by the business forecast value superimposing certain fluctuation coefficient and redundancy. This method has the following deficiencies:

- Deterministic traffic analysis method overlooked the probability statistical characteristics of network traffic as random variables.
- The value of redundant has uncertainty and is lack of quantitative estimation method.
- The region used for business forecast has no relationship with the traffic distribution.

In order to overcome these disadvantages of deterministic analysis methods, to rationally allocating network resource, this paper presents a new method of WCDMA traffic estimation. This method will export a quantitative result of traffic estimation

based on analysis of three key characteristics: traffic volatility, traffic disequilibrium at different time and in different regions.

## 2 A New Assessment Method of WCDMA Network Traffic

In order to reduce the impact of regional disequilibrium on the assessment of traffic, the existing network can be divided into certain regions. That is to merge neighboring cells with similar traffic characteristics into several continuous regions. If the divided geographical region is small enough, then we can calculate the total traffic of the system by the integral operation.

### *Geographical disequilibrium*

The operation of dividing work should be considered in accordance with cell traffic. We can extract cell-level traffic data from the OMC, and consider both voice traffic and data traffic to do the dividing work of the system.

### *Volatility*

Analysis of volatility is as follows:

- Since each user's behavior can be regarded as independent of each other, and the traffic of different users' impact on the overall system is very small, its overall impact can be seen as the Gaussian distribution, so the total system traffic as a random variable can also be seen as a Gaussian distribution.
- Fluctuations of the total system traffic can be regarded as the variance of the Gaussian distribution.
- According to statistical sample data of the system traffic in busy hour, fitting the function  $Y(t)$  to predict the trend of changes in traffic.
- According to the fluctuations of traffic, calculating the volatility factor  $\rho$ .

$\rho = \frac{Y(k) - Y(t)}{Y(t)}$ ,  $Y(k)$  is the total traffic of the day k,  $Y(t)$  is the forecast value of the day k.

- Using a series of sample values to estimate the mean and standard deviation of the total system traffic.
- Let the proposed value of the system traffic be  $P_{pro}$  and the forecast value

be  $P_{for}$ , then  $\frac{P_{pro}}{P_{for}} - 1$  satisfies the Gaussian distribution which has the mean

0 and a variance  $\sigma$ .  $P_{pro} = P_{for} \times (1 + \alpha * \sigma)$ , different  $\alpha$  corresponding to different statistical probabilities, we can get values by looking up the standard normal distribution table.

We choose a specific divided region for the analysis of traffic characteristics, in which there are 42 cells. As the calculation process of voice service and data service traffic is similar, we only show a calculation process of volatility of the voice traffic.

After choosing sample values of total traffic within the region, we derived a linear equation of the regional traffic using the data-fitting method. According

to  $\rho = \frac{Y(k) - Y(t)}{Y(t)}$ , the volatility coefficient of 10-day traffic in the region could be calculated as follows:

**Table 1.** Volatility coefficient sample values

day	$\rho = \frac{Y(k) - Y(t)}{Y(t)}$
1	0.1054121
2	-0.0887533
3	-0.1518166
4	0.0419012
5	0.1026209
6	-0.0998325
7	0.0883759
8	0.1086948
9	-0.1293744
10	0.0318747

Then we can get the mean and the standard deviation of the total traffic in this region, the mean  $\mu = 0$ , the standard deviation  $\sigma = 0.1008$ . So the proposed value and the forecast value have a relationship  $\frac{P_{pro}}{P_{for}} - 1$  which meets the Gaussian distribution, and  $P_{pro} = P_{for} \times (1 + \alpha * 0.1008)$ . We can look up the standard normal distribution table to get values according to the value of  $\alpha$ . For example, if  $\alpha = 1$ , the statistical probability is about 84.1%, that means the suggested value should be 1.1008 times to the predicted value to meet the statistical probability of 84.1%; if  $\alpha = 2$ , the suggested value should be 1.2016 times to the predicted value to meet the statistical probability of 97.7%.

*Time disequilibrium*

Analysis of time disequilibrium is as follows:

- The time disequilibrium is reflected in differences of the busy-hour traffic of the unit region in different days. As the disequilibrium coefficient is associated with the number of days to be calculated, the n-day disequilibrium coefficient can be expressed as  $\xi = \sum_i \text{Max} \left( \frac{\text{An hour's traffic of region i in n days}}{\text{System-Peak-Hour-Traffic}} - 1 \right)$ , “n” is the number of days to be calculated, “i” is the current unit region in the system.
- According to calculated values, the related function  $f(n)$  between the time disequilibrium coefficient and calculating days can be obtained.

- Using the way of curve fitting to get the function  $f(n)$ 's expression.
- Let the proposed value and the forecast value of the system traffic in  $n$  days be  $P_{pro}$  and  $P_{for}$  respectively, then we get a relationship of  $P_{pro} = P_{for} \times (1 + f(n))$ .

We select a region which has 42 cells. The  $n$ -day disequilibrium coefficient can be expressed as:

$$\xi = \sum_i \text{Max} \left( \frac{\text{An hour's traffic of cell } i \text{ in } n \text{ days}}{\text{system - peak - hour traffic in day } i} - 1 \right)$$

Let's take  $n=10$ , then the disequilibrium coefficient's value  $\xi_1, \xi_2, \dots, \xi_{10}$  increased from 0.1429324 to 0.6358721. We can see that with the increasing of  $\xi$ , the feature of time disequilibrium is becoming more obviously.

According to different values of  $\xi$ , we can get the fitting curve of time disequilibrium coefficient and days to be calculated. The curve has an analytic function of  $f(n) = 0.401 \ln(n + 2.72) - 0.384$ . As mentioned previously, the proposed value and the forecast value of the regional system traffic in 10 days are  $P_{pro}$  and  $P_{for}$  respectively, then they have a relationship of  $P_{pro} = P_{for} \times (0.401 \ln(n + 2.72) + 0.616)$ .

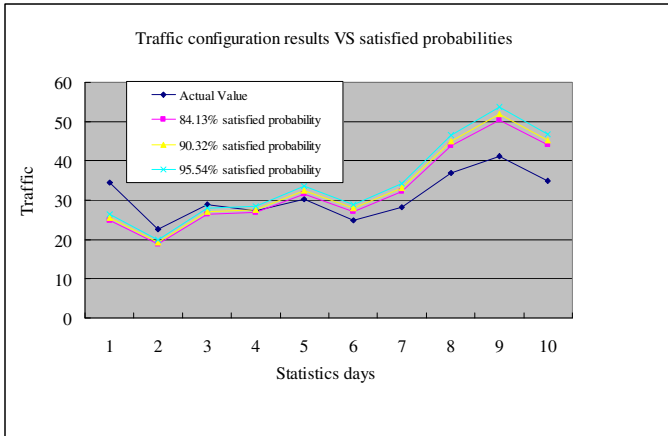
*Comprehensive assessment of traffic*

In order to get proper traffic for each divided region, we have incorporated above methods and proposed a comprehensive assessment method.

- Dividing the whole system into several regions in accordance with the geographical distribution features of traffic density.
- For one specific region, we can get the traffic assessment by combined using of volatility analysis method and time disequilibrium analysis method, that is  $P_{pro} = P_{for} \times (1 + \alpha \times \sigma) \times (1 + f(n))$ .
- So the relationship between  $P_{pro}$  and  $P_{for}$  for the current regional system should be updated to  $P_{pro} = P_{for} \times (1 + \alpha \times 0.1008) \times (0.401 \ln(n + 2.72) + 0.616)$ .
- Let the proposed value of the whole system be  $P_{pro-whole}$ , then  $P_{pro-whole} = \sum_i^m P_{pro-i} = P_{for-i} \times (1 + \alpha \times \sigma) \times (1 + f(n))$ , we can obtain the total traffic of the whole system.

So for example, there are different results of resources allocation to meet the requirements of anti-volatility with different statistics number of days. These are shown below:





**Fig. 1.** Relation between traffic configuration results and satisfied probabilities

It must be noted that, the proposed value and forecast value of both voice traffic and data traffic should be considered in practical operation to give a reasonable resource allocation scheme.

### 3 Conclusions

The result of capacity allocation in one region will affect the configuration of other regions when the total network resources are limited. Therefore, in the actual network capacity configuration work, we suggest that firstly dividing the whole area into different regions according to traffic disequilibrium characteristic and getting the allocation result at specific region based on the new method, secondly obtaining the total traffic of the whole system by aggregating allocation results of different regions.

Operators may plan the network capacity or adjust resources allocation results in different regions of the existing network according to the new method, and balance the network construction quality, network security and the cost.

### References

1. Akimaru, H., Kawashima, K.: *Teletraffic: theory and applications*. Springer, Heidelberg (1999)
2. Holma, H., Toskala, A.: *WCDMA for UMTS: HSPA Evolution and LTE*. Wiley-Blackwell (2010)

# Research on Embedded Parallel System Based on Ethernet

Li Ying<sup>1</sup> and Li Bo<sup>2</sup>

<sup>1</sup> Kunming University of Science and Technology

<sup>2</sup> Yunnan University

**Abstract.** In order to overcome the drawbacks of fixed hardware structure, difficult software development, and poor versatility of a current Multi-DSP parallel system, this paper proposes an embedded parallel system based on Ethernet. The system uses standard DSP card as a parallel processing units, which are connected as a fast Ethernet by a switch, and achieve parallel processing under scheduling of a real-time operating system. After a preliminary evaluation, the system has advantage in good versatility, flexible configuration and high parallel efficiency. It will have a very high prospect in the field of measurement and control, communication and real-time simulation system.

**Keywords:** Embedded System, Ethernet, Parallel Processing, Real-time Operation.

## 1 Introduction

Due to the specific and strong, real time, hardware and software tailored, high reliability, low power consumption, the embedded system has been widely used in all aspects of daily life, industrial control, weapons and equipments. With the development of digital technology, the scale of tasks needed to be process is getting bigger and bigger and the timeliness requirement of the tasks is also getting higher and higher in the field of measurement and control [1], communication [2], image processing [3] and real-time simulation [4] and so on. In order to satisfy these requirements, the embedded systems have been expected with huge real-time data processing calculation ability. As VLSI technology development, although, processors have a very large increase of computing power, but modern processor power consumption and heat dissipation have become the main obstacles to further improve its clock frequency. In order to overcome this obstacle and meet the high speed real time data processing requirements, introduced multi-core and parallel embedded systems technology is an inevitable trend [5].

The high bandwidth interconnection based on I/O bus is common technology solutions of parallel structure for multiprocessor embedded systems. In high performance embedded signal processing systems, DSP is mostly used as processors. Based on multi- DSP parallel processing technology, a massively parallel real-time processing system can be implemented [6]. However, some multi- DSP parallel

processing techniques can only be qualified multi- DSP systems under the certain structure and number of processors. The parallel system design is generally for specific applications, in which the structure of parallel system depends on the signal processing algorithms [7]. Since the system strictly related structures and algorithms, versatility, flexibility, and scalability of the system are poor. Moreover, the embedded system designers should not only have strong abilities of hardware design but also have strong abilities of software development, particularly the development capabilities of lower layer code. This requirement will limit the development and application of the embedded parallel system.

Another solution is using various communication interfaces of a microprocessor or extended for the microprocessor, such as UART, CAN, USB, Ethernet to achieve interconnection between multi processors. In these interconnection interfaces, Ethernet is the most widely used form of LAN construction today. TCP/IP protocol together with Ethernet has currently becoming the most popular network solution, and achieving a great success in the commercial application of computer networks. Using the Internet technology based on TCP/IP Protocol, the distributed parallel processing has been successfully achieved in the field of commercial computer, implementing interoperability of the Internet resources, including computing resource, storage resources, information resources, etc. However, the main purposes of the development of grid technology are not exactly the same as the real-time data processing requirements of embedded parallel systems. There is uncertainty in network communication based on the TCP/IP protocol, communication conflict may occur, and the communication reliability is also not meet the requirements of real-time systems.

Based on the above, this paper will discuss the feasibility of embedded parallel system constructed by using Ethernet technology. With the help of self designed information processing unit based on ADSP BF538 as the core part and Lan91C111 as the Ethernet node controller, this paper analysis and research on characteristics of embedded parallel system based on Ethernet technology, and provide feasible solutions in order to build with a high degree of versatility, flexibility and scalability of embedded parallel systems.

## **2 Architecture of Embedded Parallel Systems**

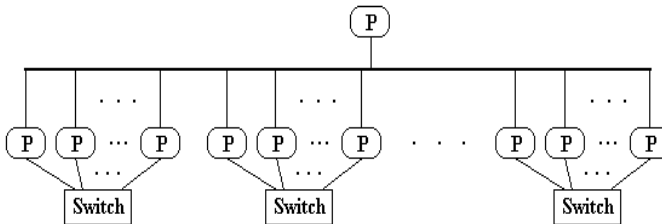
The performance of an embedded parallel information processing system depends largely on the three basic elements: parallel system architecture and information processing unit, communications and protocol between the units, monitoring and scheduling algorithm of parallel tasks of the systems, which are closely linked and mutually dependent. This paper will primarily focus on the parallel system architecture and parallel communication protocol.

Multiprocessor interconnect architecture will greatly affect the operational efficiency of a parallel system. Currently, common structures are: bus structure, linear array structure and the grid structure. All nodes in parallel bus structures share the same physical line, and only one processor is allowed to send messages at the same time. When more than one processor need to send a message at the same time, a kind of arbitration mechanism must be used to determine the order of sending messages.

Linear array column structure is the most simple of interconnected way in parallel systems, in which each node only connected with adjacent two nodes. The advantage of the linear array column structure is data transmission delay in the network can be determined due to each node has a physical chain line directly connected with other two nodes. So, the system has quite simple communication control mechanism and high real-time performance, but is not adaptation the parallel calculation need more node interconnected. In the grid-like structures, each node has a physical chain line directly connected with all other nodes in the network. A benefit of this network is more paths between nodes, so that communication blocking can be greatly reduced. But, the processing model is more complex and has higher costs.

According to the characteristics of Ethernet communication, communications of system monitoring and scheduling and data communications between nodes can be separated by using different communication interfaces to build two physically independent networks in order to get better real-time performance of the parallel system. Because the system monitoring and scheduling is global and typically has a master node (or by another node concurrently) to complete the entire system monitoring and scheduling tasks, all nodes in the system are necessary and directly connected to the primary node in order to ensure the system's real-time by selecting a bus structure to constitute a monitoring and scheduling network. Because little data flow in the monitoring and scheduling, also, a kind of TDMA time slots reserved manner can be used to ensure the dissemination of real-time monitoring and scheduling information. Other node status information with lower real time requirements can be gained by polling from host node.

In most of the parallel system, not every node requires data exchange with all other nodes. According to the demands of data exchange among nodes, therefore, the data exchanges of the entire parallel system can be done by several independent communications network. Each network uses a single switch to interconnect the nodes that requires data exchange. This architecture can not only guarantee the real-time and deterministic node data exchange, but also overcome the limitation of the node number a single switch accessed to ensure flexible and reconfigurable parallel system configurations. The architecture of the designed parallel system is showed by figure 1.



**Fig. 1.** Parallel system architecture

### 3 Ethernet Communication Protocol

High speed data transfer between nodes is the key point to achieving high performance parallel system. As advances in Ethernet technology and 1000Mbit/s and the 10Gbit/s Ethernet publication and using, Ethernet has become de facto communication standard of Internet and is also a communications technology with the highest data transfer speeds, which indicate a clear development direction for unified industry standard of real-time communication in distributed control system [8].

In the IEEE 802.3 Protocol, only two layers of PHY and MAC are defined for data communication. Higher-layer protocols are required to complete the network communications. TCP/IP Protocol is the most widely used network communication protocol, but the standard TCP/IP protocol stack is to target the business network and office Automation, which spent a lot of timer assurance on the reliability of data transmission and control of data flow and is a non-deterministic networks. Using this non-real-time and uncertainty of network protocol directly to the real-time parallel system of interconnected communications, there are obvious flaws. With the rapid development of high-speed and ultra high-speed Ethernet technology, the speed of the network interface has reached a very high level, and the actual rate of network communication protocol mainly depends on the operating efficiency of software achievement. Typically, the more protocol stack layers, the more consume CPU resources, the more spend communication, the lower operating efficiency. When Ethernet is used in an embedded parallel system, all network nodes are in the same LAN, so that IP Protocol between networks and transport layers of the TCP Protocol can be cropped and data exchange between nodes can be achieved directly based on the MAC address of network transport mechanisms.

In order to clarify this issue, we use the information processing unit based on ADSP BF538 and Lan91C111 to implement two groups of Loopback Data communication test. Each group test use three different transmission methods, one is using standard UDP in Linux operating system, the other is using Lan91 Ethernet driven code constitute of MAC1 Protocol provided by ADI company, the third is using self designed efficient Lan91 Ethernet driven code constitute of MAC2 Protocol based on parallel processing requirements. Due to the length the data portion of Ethernet frame is 46~1500 bytes, the three methods are used to send and receive maximum data frames and minimum data frames and the test results are as follows:

**Table 1.** Send and receive 1500 bytes Unit : us

Number Mode	1	2	3	4	5	6	7	8	9	10
UDP	3325	3320	3136	3451	3136	3112	3402	3188	3243	3219
MAC1	1958.4	1957.3	1958.5	1958.3	1958.2	1957.5	1958.8	1958.4	1958.2	1957.5
MAC2	1501.5	1500.5	1500.3	1500.5	1500.7	1500.3	1500.1	1500.7	1500.5	1500.6

**Table 2.** Send and receive 46 bytes

Unit : us

Number Mode	1	2	3	4	5	6	7	8	9	10
UDP	1132	1124	1122	1064	1146	1133	1167	1119	1129	1098
MAC1	125.9	125.7	126.0	125.1	125.8	125.6	125.9	126.7	125.8	126.0
MAC2	90.40	90.07	90.10	90.11	90.36	90.06	90.39	90.11	90.12	90.24

The following conclusions can be obtained by contrasting two groups of data of experiment:

- a. When sending data with the same length, much more time will be spend by using UDP communications than using MAC layer transmission.
- b. The less send and receive data, the lower UDP communication efficiency, but the situation is not evident when sending and receiving data with MAC layer directly.
- c. The communication rate with UDP is unstable, but communication rates directly in the MAC layer to send and receive data communication rates are relatively stable.
- d. According to the requirements for parallel computing, the optimized driver in MAC layer can further improve data communication rate.

Therefore, it could be concluded that direct using MAC layer to send and receive data within a local area network could not only reduce the communication overhead and enhance communication rate, especially in short frames, but also improve the certainty of communication. These characteristics have important values to meet the communication needs of embedded parallel systems and improve the real time performance of the system.

Based on optimization MAC driver code, we define several data frame formats, which is encapsulated within the data frame of Ethernet, as the basic communication protocols of the parallel system based on 100M Ethernet. Through repeated testing, the time required of sending at least 46 bytes short frame is approximately 35us, and the time required of sending of sending 1500 bytes long frames is approximately 560us. In parallel simulation applications, the amount of data exchanged between nodes is small, so that a short frame can usually be adopted to do the data transmissions. It is means that the data communication can be completed within 35us. It is a better value for coarse-grained parallel system.

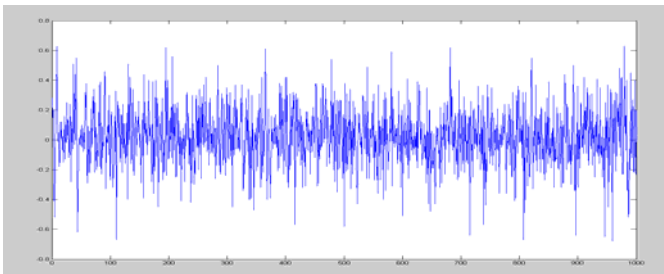
## 4 Clock Synchronization of Parallel System

In an embedded parallel system, it is different to coordinate the task of each node if there is no uniform and accurate clock. Therefore, an embedded parallel system must have a synchronous clock with high accuracy as a benchmark for realization of task scheduling in parallel and real-time control. IEEE1588 currently is the most common clock synchronization protocol for distributed network systems, which use a time synchronization technology of master-slave mode. It assumes network transmission path is symmetric. The master clock sent synchronization message cyclically to all slave clocks in the system. Using master clock as a reference, the slave calculates the

time differences with the master clock by analysis the received synchronization message and implement system clock synchronization.

In a general business network, UDP protocol can be used to implement IEEE1588 Message transmission, but the clock synchronization precision can only reach milliseconds, which is clearly not enough to meet the demand of synchronization performance for a real-time parallel system. For this reason, a parallel system clock synchronization protocol has been designed based on Ethernet MAC layer communication protocol. Since clock messages are only with little bytes of data, the system use 46 bytes shorter frame for clock synchronization communication. Specific method is that a slave node first sent a slave frame to host and record the accurate moments  $T_{s1}$  when the slave frame is sent. The host node will record the moments  $T_{m1}$  when slave frame is received. Then, host node will send a master frame to the slave node and record the accurate moments  $T_{m2}$  when the master frame is sent. The slave node will also record the moments  $T_{s2}$  when master frame is received. Finally, the host node sends a time frame with the messages of  $T_{m1}$  and  $T_{m2}$  to slave node. After received the time frame, the slave node can calculate the clock deviation of offset according to the four time values of  $T_{s1}$ ,  $T_{m1}$ ,  $T_{s2}$  and  $T_{m2}$ . The clock of slave node can be corrected based on the deviation, so that the host and slave nodes can achieve the clock synchronization. This method can simply the clock synchronization procedure, and also enhance the accuracy of clock synchronization.

In order to measure synchronization result, a hardware interrupt method was used to interrupt both host and slave boards at the same time. Both host and slave recorded the time values of their own clock, and then the slave sent a data frame with its clock value to the host. After received the data frame, the host can compare the slave clock value with its own clock value recorded and produce the synchronization error between them. In the system, the host produces measurement pulse every two second and measure synchronization error. Total 1000 data have been collected as shown on figure 2. In the figure, ordinate unit is time us, horizontal unit is the number of data.



**Fig. 2.** Errors of clock synchronization

We can see from the figure that the fluctuation range of clock error is quite less and the maximum value of the error is only 0.7us. Under the condition of no extra hardware, the synchronization accuracy of the system clock is much better for 100M Ethernet based system. The system clock can meet the basic requirements of parallel task scheduling and real-time control of the system.

## 5 Instance Testing of Parallel Computing

Parallel computing can be broadly divided into task-level and algorithm-level parallelism. The embedded parallel systems discussed in this paper are mainly used to do the task-level parallel computing. Since the purpose of this paper is to discuss the feasibility of introducing Ethernet to embedded parallel system and test their performance, an embedded parallel system with five nodes was built and all nodes were connected with a the switch. For easy to realize multitask scheduling and feature expansion of the parallel system, real-time operating system uC/OSII with open source code was utilized as the development platform. A host-slave communication mode was adopted temporarily for ease realization of task scheduling and parallel tasks were implemented by the assembly line method. In the system, the role of host was to receive external data needed to be processed, and then divided the data into several parts to send to different slaves. The hose will coordinate the tasks of all salves and control the system to complete the data processing in the shortest amount of time or within a specified time. The slaves were in charge of receiving data of the host, did the data processing tasks in accordance with the appropriate algorithms and returned the result to the host.

Solving differential equations with a computer has the characteristics of smaller number of input data but longer operation time. So, it is a better instance applied to the designed embedded parallel system. Therefore, the embedded parallel system is used as a real time simulation system by solving a series of differential equations. The instance can both ensure the complexity of the operation and easily test the performance of the parallel system compared with stand-alone serial operation.

The detail experimental process is that the host first assigned the prepared initial value of the differential equation to Slave1. Slave1 solved the differential equation and sent the results to the host. After received the results of Slave1, the host immediately transferred the results to Slave2 as initial values of Slave2. When Slave2 began to solve the differential equation, the host sent the initial value required for next step of operations to Slave1 and Slave1 began new operations once again. After completed its operations, Slave2 returned results to the host and the host transferred the received results to Slave3. If Slave1 completed the second operation, while, the host transferred the results to Slave2, and then sent the initial value of the third operation to Slave1. When Slave4 completed the operation, the results for solving the differential equation were eventually returned to the host. Please reference the figure 3 of the stream-line diagram.

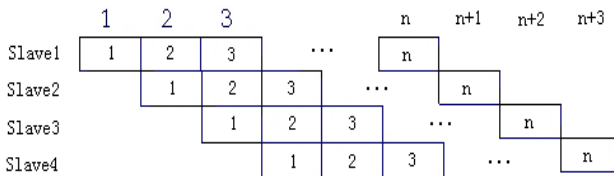


Fig. 3. Stream-line diagram



Different numbers of steps are set to test the system in here. The expression of differential equation is:  $f(x, y) = y - 2x/y$  and the initial value for  $x=0, y=1$ . The algorithm of solving the differential equation is based on the fourth-order Runge–Kutta method and calculate 128 , 256 , 512 , 1024 step numbers with step length of 0.01. The parallel computing system should minimize the errors of all node operations. Otherwise, it will affect the overall performance of the system. Due to four slaves used for the operation in the system, the step number of each slave can be defined as one-fourth of the total number of steps. In the experiments, the host is timing the process of operations and determining the finished time. Due to the designed system is a stream-line parallel computing systems, the period of time for solving the differential equation should measure multiple (greater than line progression) time other than once in order to evaluate the performance of the parallel system.

According to the design of above experiment, the DSP parallel system and single DSP board are used to do the same experiment, in which the fourth-order Runge–Kutta is utilized to solve the same differential equation in many different steps. The overall spending time was measured by an on board timer of the host, and experiment results are showed as follow:

**Table 3.** Spending time of 10 operations Unit : ms

Steps	128	256	512	1024
Parallel computing time	0.121	0.219	0.412	0.771
Serial computing time	0.345	0.660	1.280	2.437

**Table 4.** Spending time of 50 operations Unit : ms

Steps	128	256	512	1024
Parallel computing time	0.576	1.031	1.935	3.679
Serial computing time	1.720	3.284	6.436	12.244

According to the above experiment data, two important performance indicators, speed-up ratio and parallel efficiency, of the parallel system were calculated. Setting the serial execution time as  $T_s$  and parallel execution time of  $q$  processors as  $T_p(q)$  , the s speed-up ratio is:

$$S_p(q) = \frac{T_s}{T_p(q)}$$

and parallel efficiency of the algorithm is:

$$E_p(q) = \frac{S_p(q)}{q}$$

**Table 5.** Speed up ratio

Steps	128	256	512	1024
10 time calculations	2.851	3.014	3.107	3.161
50 time calculations	2.986	3.242	3.326	3.328

**Table 6.** Parallel efficiency

Steps	128	256	512	1024
10 time calculations	0.570	0.603	0.621	0.632
50 time calculations	0.597	0.648	0.665	0.666

By analyzing above experimental data and the performance indicators of the system, it can be concluded that the time complexity of the designed embedded parallel systems is greatly reduced compared with serial operation. With the increase of the step number and calculation of Runge – Kutta method, speed-up ratio and parallel efficiency of the system will be improved, which illustrates that the parallel computing system is more suitable for the occasions with a large amount of processing but small communication traffic, such as coarse-grain parallel algorithms. The reason of the result is due to the proportion of the time used for data traffic during the entire operation declined with the growth of calculation time for each processor in the system, so that speed-up ratio and parallel efficiency of the system can be improved in different degrees.

## 6 Conclusions

This paper describes a basic framework of the multi DSP parallel system based on Ethernet. Through the application of a specific instance, the actual performance of the parallel system has been tested. The experiment result shows that the system has better prospects in the field of coarse-grained parallel computing with a large amount of processing but small communication traffic. Compared with traditional multi- DSP parallel systems, the Ethernet based multi- DSP embedded parallel system has several advantages, such as low development difficulty, easy standardized hardware and software architecture, flexible system configuration, easy extension and so on. With the popularizing and application of 1000Mbit/s and the 10Gbit/s Ethernet, the data exchange rate between nodes of the parallel system will has great improvement.

The bottleneck limited efficiency improvement of the parallel system will gradually be overcome. The application range of the parallel framework will become wider and wider. A low cost framework of standardization embedded parallel system has been provided for real-time measurement and control system, and communication system and real-time simulation system.

Since this paper is written according to the prophase study, only the basic framework and basic performance of parallel systems are described in this paper. The research of how to extend the network bandwidth to achieve multi-channel parallel communication with a switch and how to improve real-time operating system to implement multi tasks parallel scheduling will be discussed in later papers.

## References

1. Gamatie, A., Gautier, T.: The Signal Synchronous Multiclock Approach to the Design of Distributed Embedded Systems. *IEEE Transactions on Parallel and Distributed Systems* 21(5), 641–657 (2010)
2. Jin, H.-W., Yun, Y.-J., Jang, H.-C.: TCP/IP Performance Near I/O Bus Bandwidth on Multi-Core Systems: 10-Gigabit Ethernet vs. Multi-Port Gigabit Ethernet. In: *Parallel Processing - Workshops, ICPP-W 2008, International Conference on Portland* (2008)
3. Hu, J., Zhang, T., Jiang, H.: New multi-DSP parallel computing architecture for real-time image processing. *Journal of Systems Engineering and Electronics* 17(4), 883–889 (2006)
4. Goyal, S., Ledwich, G., Ghosh, A.: Power Network in Loop: A Paradigm for Real-Time Simulation and Hardware Testing. *IEEE Transactions on Power Delivery* 25(2), 1083–1092 (2010)
5. Paulin, P.G., Pilkington, C.: Parallel programming models for a multiprocessor SoC platform applied to net—working and Multimedia. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 33(7), 667–680 (2008)
6. Wang, J., Zhang, Y., Li, W.: High performance DSP method for interconnecting. *Electronic devices*, 1943–1946 (2008)
7. Assayad, I.: A platform-based design framework for joint SW/HW multiprocessor systems design. *Journal of Systems Architecture* 55(7-9), 409–420 (2009)
8. Cenac, G., Senoa, L., Valenzano, A., Vittoria, S.: Performance Analysis of Ethernet Powerlink Networks for Distributed Control and Automation Systems. *Journal of Computer Standards & Interfaces*, Article in Press, Accepted Manuscript - Note to users (2008)

# A PMAC Motor Drive with a Minimum Number of Sensors

Li Ying<sup>1</sup> and Nesimi Ertugrul<sup>2</sup>

<sup>1</sup> Kunming University of Science and Technology

<sup>2</sup> The University of Adelaide

**Abstract.** This paper proposes and implements a novel control technique for a PMAC motor drive with minimum a number of sensors. In the control scheme, the required three-phase voltages and currents are reconstructed from DC link measurements, and then the reconstructed voltages and currents are used to estimate rotor position with the robust position estimation algorithm. The estimated rotor position and the reconstructed phase currents can be used as feedback signals to implement the closed loop control in real time. The experiments demonstrate the feasibility of the implementation of the PMAC motor control with only two sensors and the motor drive has reasonable performances for some low cost applications.

**Keywords:** Position sensorless, Current Reconstruction, Minimum Sensors, PMAC Motors.

## 1 Introduction

A PMAC motor has permanent magnets mounted on its rotor and polyphase excitation windings inserted in the stator that is similar to a conventional induction motor. In order to produce a constant torque, the winding excitation currents should be synchronised with the back-EMF voltages. Therefore, the most fundamental requirement in the PMAC motor control is to know the rotor position information, which is required to generate the excitation currents and to implement self-synchronous control. In addition, a speed feedback signal may be necessary in many applications to implement the closed loop speed control. Furthermore, in some applications, such as PMAC servo motor drives, the torque control is required to achieve a high dynamic performance, which can easily be done by mapping it to the current control that force the actual phase currents to track the reference currents. Therefore, a high performance PMAC motor drive requires a number of sensors to obtain the necessary feedback signals, the rotor position, the rotor speed, and the three-phase currents.

In a practical motor drive, the rotor speed data can be derived from the rotor position information, so that the speed sensor can be eliminated. In addition, if the motor is star-connected, the number of current sensors can be reduced to two since the third phase current can be derived easily using the other two-phase currents.

Therefore, it may be sufficient to use only three sensors (one position sensor, two phase current sensors) to implement a high performance motor drive.

Traditionally, the rotor position information is obtained directly from the motor shaft using some form of shaft-mounted position sensor, but the sensor often is one of the most expensive and fragile components in the entire motor drive system. Elimination of the shaft-mounted position sensors is very desirable objective in many applications by using a position sensorless technique. However, in order to estimate the rotor position accurately, most of the successful techniques require accurate measurement of the three-phase currents and the voltages in the motor drive, which means that five sensors are still required to implement the position sensorless PMAC motor drive.

In order to isolate high-voltage inverter circuit from low-voltage control circuit, Hall-Effect current sensors and isolated amplifiers are usually used to measure currents and voltages in a practical motor drive. These voltage and current sensors are expensive for some low-cost applications. Therefore, employing a minimum number of sensors, which also increase the reliability of the system, and reducing cost are desirable in some commercial applications.

This paper focuses on the feasibility of minimizing the sensor number to implement a PMAC motor control. Based on the robust position sensorless technique [1] and the reconstruction methods of phase currents and voltages [2], a real time PMAC motor drive system with a minimum number of sensors was developed. The motor drive uses two sensors to measure DC link current and voltage of the inverter respectively. If the input voltage of the inverter remains constant, the voltage sensor can be also eliminated and only one current sensor is required in the motor drive system. Real time experiment results are provided in this paper to show the performances of the single sensor motor driver system. The differences between sinusoidal and trapezoidal PMAC motors are also discussed in this paper.

## **2 Minimizing Current and Voltage Sensors**

In most PMAC motor applications, the knowledge of the phase currents and the voltages is required to implement the closed-loop current control and to accommodate inverter and drive protection. In the position sensorless motor drive system, the phase current and voltage signals are also required to estimate the rotor position. As stated in [2], in order to minimize the number of sensors, the three-phase currents and voltages of the PMAC motor can be reconstructed using the DC link measurements and the switch states of the inverter.

### **A Phase Voltage Reconstruction**

A three-phase inverter consists of six power switches and six freewheeling diodes. Two devices (power switch and freewheeling diode) in one leg could be regarded as a

normal mechanical switch, which can conduct current bi-directionally, although the conduction conditions in either direction are different. Since two switches in one phase can be open but not closed at the same time, a state variable,  $S_p$  ( $p = a, b, c$ ) with three states can be used to represent the two switches: when the upper leg switch of phase P is closed,  $S_p=1$ ; when the lower leg switch of phase p is closed,  $S_p=-1$ ; and when two switches are all open,  $S_p=0$ .

With the help of switch state variables, three-phase voltages of a PMAC motor drive can be reconstructed from DC link voltage. The unified expression of the phase voltage reconstruction can be expressed as follow [2],

$$v_p = \frac{1}{K} \left[ \frac{(K \cdot S_p - S_a - S_b - S_c) \cdot V_{dc} / 2}{+ (|S_a| \cdot e_1 + |S_b| \cdot e_2 + |S_c| \cdot e_3)} \right] \tag{1}$$

In the above expression, the switch states ( $S_a, S_b, S_c$ ) depend on the switching signals of the inverter and the directions of the phase currents, and the DC link voltage  $V_{dc}$  can easily be measured. However, the three-phase back EMF voltages ( $e_a, e_b, e_c$ ) cannot be measured directly but can be generated from the rotor position and the speed signals.

If back EMF voltages of a PMAC motor are sinusoidal, the sum of the three-phase back EMF voltages is equal to zero. When all three phases are excited with the currents, such as three-phase sinusoidal currents, there are only eight possible switch states, which make the phase voltage reconstruction simple:

$$v_p = \left[ S_p - \frac{1}{3}(S_a + S_b + S_c) \right] \cdot V_{dc} / 2 = \frac{V_{dc}}{6} (3 \cdot S_p - S_a - S_b - S_c) \tag{2}$$

This expression indicates that the phase voltages of the sinusoidal PMAC motor only depend on the DC link voltage and the switch states.

**B Phase Current Reconstruction**

When the motor is powered from an inverter, the three-phase currents of the motor are closely related to the DC link current of the inverter. Based on this relationship, the three-phase currents can be reconstructed from the DC link measurements with an adaptive state observer [2]. The method uses the mathematic model of the PMAC motor drive as a state observer to estimate three-phase currents. Using estimated phase currents, a DC link current is constructed as the output of the observer. Then, assuming three-phase symmetry in the PMAC motor, the observer output is compared with the measured DC link currents and produces some compensation on the DC link voltage that is used to reconstruct the three-phase voltages in the current observer. Such compensation can correct the estimated three-phase currents and force them to follow the actual winding currents under all operating conditions. Figure 1 shows block diagram of the current observer.

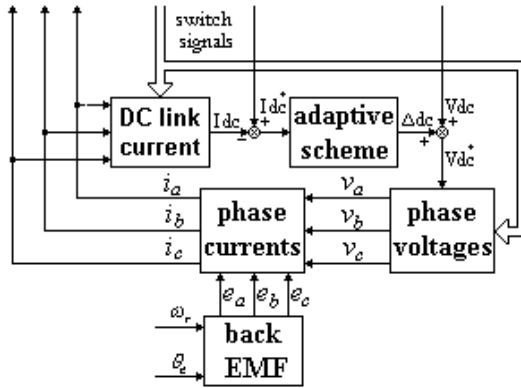


Fig. 1. Block diagram of current observer

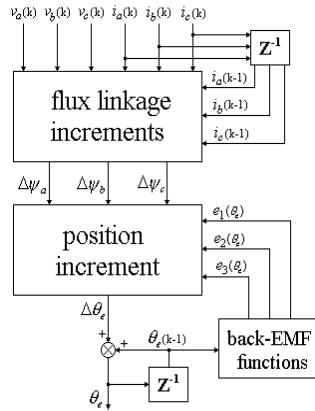
It should be noted here that the current reconstruction technique is based on the current observer that requires the back EMF voltages, which cannot be measured but were generated using the estimated rotor position and the speed. This means that the reconstructed phase currents and the estimated rotor position are interrelated. However, the back EMF voltages play a minor role in the current estimation as can be seen from the mathematic model. Moreover, the current observer is a closed loop adaptive observer. The modified DC link voltage can compensate the errors in the estimated back EMF voltages. Unlike the phase voltage reconstruction in the trapezoidal PMAC motor, the reconstructed phase currents can still be used to estimate the rotor position and the speed even though their reconstruction is related to the back EMF voltages.

The experiments results given in [2] have demonstrated that the current observer can reconstruct the three-phase currents of the PMAC motor accurately using the DC link measurements and the switching signals of the inverter. As also demonstrated, the reconstructed phase currents have smaller static errors and faster dynamic responses and can be used as feedback signals to implement the closed loop current control in the PMAC motor drive.

### 3 Elimination Rotor Position Sensor

In recent years, a variety of position sensorless techniques have been proposed for the self-synchronous control of a PMAC motor drive. These techniques can be broadly classified as three typical kinds.

The position sensorless technique detailed in [1] has the robustness with respect to these differences. Therefore, the integration of the two techniques, the reconstruction of phase voltages and currents and the robust rotor position estimation, can minimize the number of sensors used in a PMAC motor drive. Figure 2 shows block diagram of the robust position estimator.



**Fig. 2.** Block diagram of the robust position estimator

In this new scheme, the calculation of the flux linkage is replaced with the direct calculation of the flux linkage increment using the phase current and voltage signals, that is,

$$\begin{aligned}
 \Delta\psi_a &= (v_a - R i_a(k)) \cdot \Delta t - L(i_a(k) - i_a(k-1)) \\
 \Delta\psi_b &= (v_b - R i_b(k)) \cdot \Delta t - L(i_b(k) - i_b(k-1)) \\
 \Delta\psi_c &= (v_c - R i_c(k)) \cdot \Delta t - L(i_c(k) - i_c(k-1))
 \end{aligned}
 \tag{4}$$

It can be seen that the calculation above has no integration, which is required in the calculation of flux linkage and is the main source of the estimation error in the previous techniques. In addition, the back-EMF voltages are disassembled as the product of angular speed and the unit back-EMF functions that become independent of the angular speed. Using the increments of flux linkage and the back-EMF functions of the motor, a single rotor position increment can be estimated with following equation:

$$\Delta\theta_e = \frac{N_p}{k_e} \cdot \frac{\Delta\psi_a e_2(\theta_e) + \Delta\psi_b e_3(\theta_e) + \Delta\psi_c e_1(\theta_e)}{e_1(\theta_e) e_2(\theta_e) + e_2(\theta_e) e_3(\theta_e) + e_3(\theta_e) e_1(\theta_e)}
 \tag{5}$$

Finally, the corresponding rotor position information can be calculated by integrating the rotor position increment. It should be noted that the back-EMF functions used in the calculation are generated from the estimated rotor position that forms an internal closed loop in such estimation scheme. In this way, the integration existed the flux linkage calculations is transferred to the position calculation, that is, from outside to inside of the closed loop. Similar to a phase-locked loop (PLL), the Equation 5 has a kind of auto-correctional capability that can correct position drift caused by the motor parameter deviations and/or the measurement inaccuracies via the internal closed loop.



### 4 PMAC Motor Drive with Minimum Number of Sensors

By integrating the two techniques together, a novel PMAC motor control scheme was formed as the block diagram shown on Figure 3. The PMAC motor drive system uses only two sensors to measure DC link current and voltage respectively. Moreover, if the DC link voltage remains constant (the inverter is connected to utility power lines), a further reduction of the sensors is possible by eliminating the voltage sensor.

In this scheme, the three-phase voltages and currents of the motor are reconstructed from the DC link measurements and switching signals of the inverter at first. Then, the reconstructed voltages and currents are used to estimate rotor position with the robust position estimation algorithm.

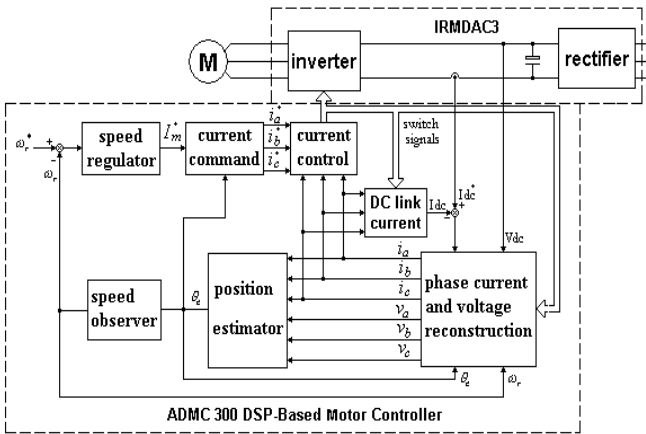


Fig. 3. A PMAC motor drive with minimum number of sensors

Furthermore, it should be emphasised here that the phase voltages should be reconstructed twice in the integrated motor drive. First, in the current observer, the measured DC link voltage must be compensated by the adaptive scheme using the DC link current error and then can be used to reconstruct the modified phase voltages, which are used in the estimation of the phase currents. Second, the actual measured DC link voltage should be directly used to reconstruct the phase voltages that are required in the position estimation algorithm.

Consequently, the speed observer uses the rotor position to estimate the angular speed signals, which is then used as the speed feedback to implement the closed loop speed control. The output of the speed regulator is defined as the amplitude of the current command. Together with the estimated rotor position signals, the amplitude is used to generate the required current reference signals, and the reconstructed phase currents are used as the current feedback signals.

To implement the control scheme in real time, an ADCM 300 DSP-based Motor Controller was used, but only two A/D converter channels on the controller was utilized to measure DC link current and voltage of the inverter respectively. The software of the implementation was separated into two interruption service routines at different interruption frequencies. One of the service routines, which have the same

interruption frequency as the PWM frequency, performs the speed regulation, the current control and the PWM output. The algorithm of this service routine is same as the conventional control except the phase current signals, the rotor position and the speed signals were not obtained via the A/D converters and but linked to the second service routine.

The second service routine performs the reconstruction of the three-phase currents and voltages, the estimation of the rotor position and the rotor speed observer. As detailed in [2], the phase current reconstruction directly depends on the accuracy of the DC link current measurement. Therefore, the sample rate of A/D converter should be as high as possible in the real system. The highest possible sample rate of the A/D converter of the ADMC 300 DSP-based motor controller is 31.5KHz. Based on this hardware limitation, the experiences indicate that the PWM frequency could be selected as 1302Hz in order to provide reasonable accuracy in the measurement of the DC link current. This means that the interruption frequency of the reconstruction and the estimation service routine is 24 times of the PWM frequency in the real time system, which allowed us to implement the PMAC motor control with two sensors.

According to the structure of the software implementation, the reconstruction and the estimation service routine perform 24 times in each PWM cycle, but the reconstructed phase current, the estimated rotor position and the speed values are updated once per PWM cycle. Therefore, 24 reconstructed and estimated values within one cycle can be averaged or filtered, which may produce better results by reducing the fluctuations in the reconstructed three-phase currents and the random position estimation errors.

## 5 Real Time Experiment Results

The PM motor used in the experiments has sinusoidal back EMF and its parameters are given in Appendix.

In the real time implementation of the PMAC motor control utilising two sensors can start directly from zero speed in real time without accommodating any special starting strategy. At beginning of the starting, the reconstruction algorithms predict the phase voltages and currents using the measured DC link current and the voltage, and the position estimation algorithm uses the reconstructed values to estimate the rotor position using a random initial value. Although this may generate wrong estimation, within one electrical cycle, the reconstructed phase currents and the estimated rotor position can converge to the actual values and can be used as the correct feedback signals in the closed loop system. Due to the large errors of the reconstructed phase currents and the estimated rotor position at beginning, the dynamic performance of the starting procedure of the motor drive is not as good as expected. However, it was observed during the tests that the motor drive could always start smoothly with this starting approach.

The first test result shown in Figure 4 illustrates the operation of the two sensor motor drive at a steady-state speed 107rpm (about 50Hz). In the figure, the measured DC link current, the reconstructed three-phase currents, and the estimated rotor position are given. The test is repeated at a much lower operating speed of 21.4rpm

(about 10Hz) and the results are presented in Figure 5. As demonstrated, the waveforms of the reconstructed three-phase currents and the rotor position are used as the feedback signals in the drive and the closed loop control was achieved in the motor drive.

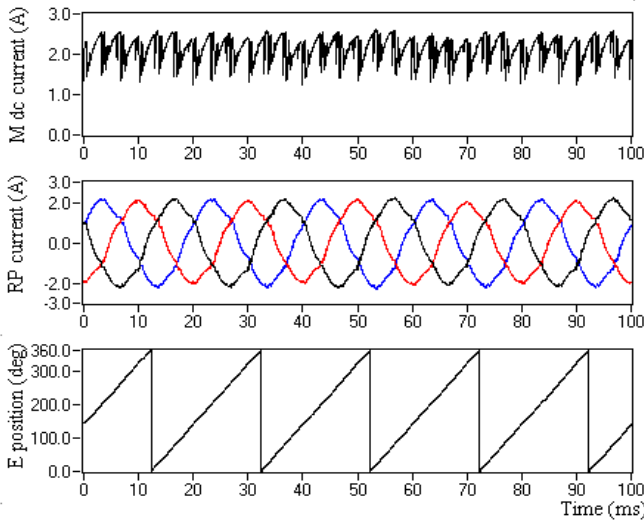


Fig. 4. Steady state experiment results of the motor drive running at 50Hz

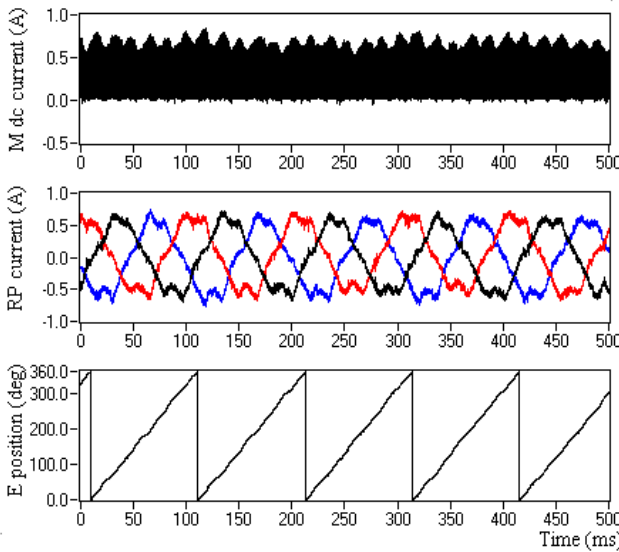
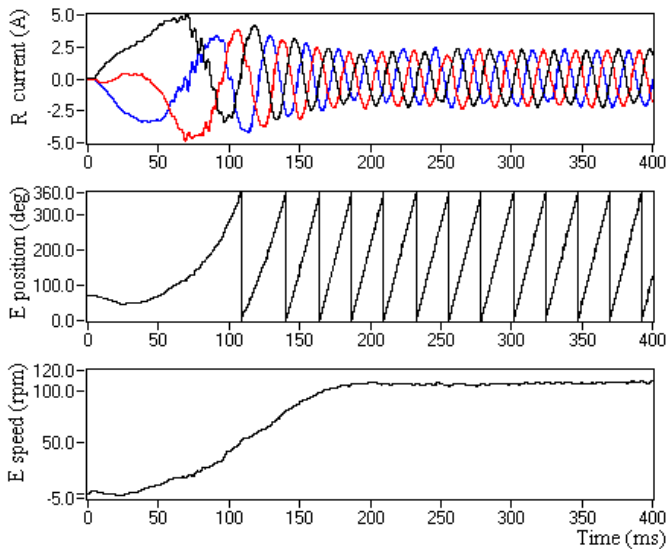


Fig. 5. Steady state experiment results of the motor drive running at 10Hz

Figure 6 shows an ultimate test result to demonstrate the dynamic performance of the motor drive operating in closed-loop and using only two sensors. The transient waveforms of the three-phase currents, the estimated rotor position and the speed are all given in the figure. In this test, the motor drive starts from standstill, accelerates and reaches to a steady-state speed of 107 rpm within 200ms. The results indicate that the motor drive can start directly from zero speed and also demonstrates a reasonable dynamic performance when uses the reconstructed phase currents in the current control loop and the estimated rotor position to generate the reference currents.



**Fig. 6.** Dynamic experiment results of the motor drive

In comparison with the results detailed in [1,2], it was found that the performances of the motor drive are not very accurate and the operation speed range is limited. The primary reason of such weaknesses is the limitation of the available hardware as described earlier. To improve the performance of the two-sensor motor drive, a higher speed DSP with higher performance A/D converters should be used, such as ADMC 501. In addition, the speed observer, the speed regulator and the current controller in the drive system should be carefully redesigned.

## 6 Conclusions

The paper discussed a possible solution for the sensor reduction in the PMAC motor drives. In the proposed control scheme, the two techniques, phase current and voltage

reconstructions and robust position sensorless, are combined together to implement the PMAC motor control with only two sensors, sensing the DC link current and voltage respectively. If the DC link voltage remains constant, the voltage sensor can be eliminated also and the motor drive requires only one current sensor.

In the motor drive, the required three-phase voltages and currents are reconstructed from DC link measurements, and then the reconstructed voltages and currents are used to estimate rotor position with the robust position estimation algorithm. When the motor is excited with sinusoidal currents, the estimated rotor position and the reconstructed phase currents can be used as feedback signals to implement the closed loop control in real time. The experiments demonstrate the feasibility of the implementation of the PMAC motor control with only two sensors and the motor drive has reasonable performances for some low cost applications. When the motor is excited with rectangular currents, however, the motor drive cannot operate properly and motor system may vibrate when operation conditions are changed. This implies that the motor control technique could not be suitable for a trapezoidal PMAC motor.

Due to the limitation of the available hardware and low PWM frequency, the performances of the motor drive with only two sensors are not satisfied. It was suggested that to improve the performance of the motor drive, a higher speed DSP with high performance A/D converter should be accommodated in the controller, such as ADMC 501.

It should be emphasized here that this research focused on the sensor minimization techniques only for the closed-loop control of the PMAC motor drive. To protect the motor drive against the potential faults, additional sensors may still be needed, but this is not covered in this study.

## References

1. Ying, L., Ertugrul, N.: A Novel, Robust DSP-Based Indirect Rotor Position Estimation for Permanent Magnet AC Motors without Rotor Saliency. *IEEE Transactions on Power Electronics* 18(2), 539–546 (2003)
2. Ying, L., Ertugrul, N.: An Observer-Based Three-Phase Current Reconstruction using DC Link Measurements in PMAC Motors. In: *IPEMC 2006* (2006)
3. Iizuka, K., Uzuhashi, H., et al.: Microcomputer Control for Sensorless Brushless Motor. *IEEE Transactions on Industry Applications*, 595–601 (May/June 1985)
4. Rusong, W., Slemon, G.R.: A Permanent Magnet Motor Drive Without a Shaft Sensor. In: *Conference Record of IEEE IAS Annual Meeting*, pp. 553–558 (1990)
5. Kulkarni, A.B., Ehsani, M.: A Novel position Sensor Elimination Technique for the Interior Permanent Magnet Synchronous Motor Drive. *IEEE Transactions on Industry Applications* 28(1), 144–150 (1992)

## Appendix

The parameters of the experiment motor:

Back EMF constant, $k_e$	3.785 V/rad/s
Number of poles pair, $N_p$	28
Winding resistance, $R$	6.4 $\Omega$
Equivalent winding inductance, $L$	32.8 mH

# The Integrated Concept of the Ground Segment for Space Internet

Jian-ping Liu<sup>1</sup>, Chun-yan Luo<sup>2</sup>, Xiao-ying Li<sup>2</sup>, and Fan Yang<sup>2</sup>

<sup>1</sup> State Key Laboratory of Astronautic Dynamics, XSCC, Xi'an 710043, China

<sup>2</sup> XSCC, Xi'an 710043, China

ljpnudt@sina.com, snowdogdlc@163.com, season01@sina.com,  
Lilyang1014@163.com

**Abstract.** With the rapid development of internet and information technology as well as the increasing complexity of space missions, a strong demand to extend the internet to space and to establish an integrated space-ground network system came into being. To develop the cooperating capabilities of satellite-ground integrity and to improve the inter-operational capabilities of different systems of ground segment, the importance of integrated research for ground segment is presented in the paper firstly, then the integrated concept of the ground segment for space internet was proposed, and some key problems that needs to be solved are analyzed. Finally, suggestions are given regarding the steps of development.

**Keywords:** space internet, system integration, inter-operation.

## 1 Introduction

With the rapid development of internet and information technology as well as the increasing complexity of space missions, a strong demand to extend the internet to space and to establish an integrated space-ground network system came into being. The nodes of the network have already extended into the LEO, GEO, the moon, the solar system to construct the Interplanetary Internet in the United States [1]. Combined with the existed questions of Chinese space missions and the demand of future space missions, Academician Shen Rongjun proposed the overall objective, the composition, the network system architecture, and the network protocols of integrated space-ground network system [2]. Space internet comprises space segment and ground segment, which formed the integrated internet network through space-ground linking. The ground segment includes all the aerospace facilities, such as data receiving network, including civil and military satellite ground receiver, space surveillance network, astronomical observation network, USB and UCB ground-based TT&C network, deep-space TT&C network, relay satellite ground system, payload TT&C network, users business application network, on-orbit diagnosis and maintenance system of development units, and the other system or network related to the space missions. To develop the cooperating capabilities of satellite-ground integrity, it needs to realize such capabilities as united management, high-efficiency

sharing, integrated operation and complex utilizing, which are the capabilities of system integrated capabilities of ground segment. Therefore, the paper discusses the integrated problems of ground segment of space internet.

The research on the integrated problems of ground segment abroad is in the phase of concept. Aiming at the condition of “one space mission, one mission center”, NASA proposed GMSEC configuration framework in 2001 to reduce the development cost and improve the time of integration. At present, five space missions have already completed. This framework will be extended from mission center to the overall ground system related to the space missions of NASA [3]. For current configuration of stovepipe systems, a compatible ground control framework was proposed in 2008 by U.S. SMC satellite control and network system group [4]. Currently, the research work of the first stage has been finished, and the research work of the second stage is being performed. The European space awareness data processing system based on the service framework was proposed to meet the need of European system [5]. Seeing from the ground segment of China’s space missions, each system of the ground segment has high inter-operational capabilities. And the inter-systems have realized data-transforming. However, the inter-operational questions between the systems are obvious. Meanwhile, it mainly depends on manual scheduling and lacks of the united management and integration to the ground segment. Nevertheless, the research on this aspect hasn’t developed yet.

## **2 The Importance of the Integrated Research on Ground Segment of Space Internet**

### **1) The requirement of the development of Space TT&C Technology**

From the trend of space TT&C system in the world, ground-based TT&C system has experienced from separated control system to unified carrier system. The performance of TT&C network is more complicated than before: from the performance of tracking and control to the performance of data-acquiring which comprises such tasks as tracking, telemetry, telecomm and, voice and data transferring, and as so on. From ground to space, the higher requirement of the capabilities of tracking, data transferring and communication are brought forward with the development of satellite application. Space-based TT&C system represented by TDRSS is the best approach to solve the problems of overall orbit tracking, high-speed data transferring and multiple target tracking. The new concept and trend of space technology proposed by the U.S. are usually ahead of any other countries in the world. The systems of space communication and navigation proposed by the U.S. will be the represented technology in the space domain. At present, China is at the development of space system construction and space technology. With the constructing of space foundation infrastructure, and the management of on-orbit satellites, as well as the high reliability, high coverage, high accuracy, high speed TT&C requirements of manned spaceflight and space station, the new urgent requirement of China’s space TT&C system which depends mainly on ground-based TT&C network would be presented. So, it is of great importance for the development of China’s space TT&C technology to establish an Integrated Space-Ground TT&C



System and improve the integrated ability of utilizing TT&C resource by learning from the experiences of space TT&C system abroad.

2) The requirement for improving the interoperability of space TT&C system.

According to the definition from IEEE standard dictionary, Interoperability is the ability to exchange and using the exchanged information across different systems. From this definition, Interoperability contains two capabilities: the capability of exchange information and the capability of using information. These two capabilities are mainly reflected on the reliability, security, scalability. In the view of the interoperability of space TT&C in our country, the development of interoperability in different subsystems is unbalanced. For example, the interoperability among active launch system of space task, rocket launch TT&C system and satellite control center is comparatively good. However, the interoperability between satellite control center and the applied or scientific task center, the interoperability between satellite control center and relay system, the interoperability between the applied or scientific task center and the relay system, and the interoperability between satellite control center and space target monitor system are not so good. Therefore, in order to improve the insufficiency of the interoperability of space TT&C in our country, with the international space task cooperation getting more and more deeply, realizing standard mechanism for inter-operate across different receiving and launching systems need us to consider space TT&C synthetically. The typical representative of international cooperation in space TT&C is “Peace” space station task. The USA, Europe, Japan, Russia used the uniform and open standard, which realized inter-operate.

3) The effective way to realize rapid response of space TT&C system.

Rapid response refers to getting launch prepared in a few hours after receiving command and executing military campaign task in a few hours after getting to orbit. Space TT&C system contains Rocket launch TT&C system, spacecraft TT&C system, applied task TT&C system etc, which are the basic systems to realize rapid response in military campaign. Rapid response not only requires each system to have rapid internal response, it more importantly requires cooperated rapid response between each system. As long as each system cooperates with each other smoothly, rapid response in military campaign task could be realized. At present, research on rapid response of space TT&C in our country is mainly focused on internal rapid response of each system, but the cooperated rapid response cross systems is short of deep research. However, the purpose of research on integrated space-ground TT&C system is to solve the cooperating problems in different systems, using information technology and integrated task planning technology to make each system cooperate with each other smoothly for achieving the common target timely Therefore, the research on ground segment integration is the effective way to realize cooperated rapid response of space TT&C in our country.

### **3 The Integrated Concept of the Ground Segment**

The integrated architecture of ground segment is displayed in Fig 1. The integrated system contains six subsystems: rocket launch TT&C system, ground-based TT&C



parts: systems network management platform, systems information interaction platform, systems operation sharing platform, and integrated mission planning platform. The four parts are from low layer to high layer, and low layer provides service to high layer.

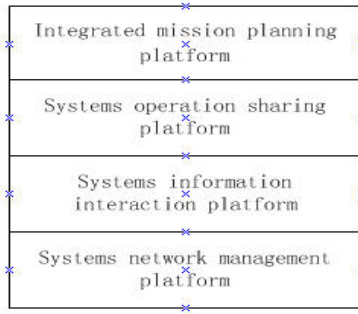


Fig. 2. Framework of platform layer

Systems network management platform belongs to network layer, performs the network monitoring function, including configuration management, failure management, fault management, TT&C resource management, and security management, and provides service to systems information interaction platform. Systems information interaction platform belongs to information transport layer, performs establishment and management of information, standardizes the information interaction activity, and provides service to systems operation sharing platform. Systems operation sharing platform belongs to operation management layer, performs the cooperating management function, standardizes the end-to-end cooperation, and provides service to the integrated mission planning platform. Integrated mission planning platform belongs to mission planning layer, performs the mission planning and scheduling function of each subsystem in platform, and coordinates the operation of each system orderly.

#### 4 The Key Problems Need to Be Solved

The ground segment of space network involves many domains, departments, software and hardware systems, heterogeneous environments, application functions. How to make them seamless and efficient work is the key problem needs to be solved. The challenges include the following four aspects:

1) The problem of integrated architecture

The integrated architecture mainly includes two problems. The first is the structure questions of inter-connection, inter-communication and inter-operation among the six systems shown before, and between each system and management center. The structure of inter-connection is topological, the structure of inter-communication is communication structure, and the structure of inter-operation is information exchange structure between end-to-end application software. Here, the precondition is inter-connection, the foundation is inter-communication, and the highest requirement or

ultimate goal is inter-operation. The interoperability of management and technology between the two different systems is the main factor which influence inter-operation, including the different development ideas and construction level brought by unified management, the different system structure, the different levels of technology, the different infrastructure, the different standard such as communication protocol or information format and so on. The second is the integrated management structure inside the system management center, including composition, function, operation mode, interface and so on.

#### 2) The problem of information sharing and exchange

The foundation of coordinate operation is information sharing and exchange, which mainly displays in two aspects. One aspect is information sharing and exchange between different systems. Because of long-term independent development of each system, its efficiency and application have been improved. But each system is disconnected because of different database, development techniques, business logic, and there is no unified standard on database design, which results to isolated information island. All data sources are heteroid that blocks information sharing and exchange. Another aspect is information sharing and exchange between management center and each system. Therefore, it is required that the standard description and uniform data format is established based on analyzing information format.

#### 3) The problem of coordinate operation across systems

In the same system, operation has some characteristics such as close, stable and controllable etc. Each system has mature flow management technology and high automation. But coordinate operation across systems has some characteristics such as applied difference, autonomy and complexity. An ideal way is to use the same system structure, software, message, protocol, safety and operation concept. But this way is unrealistic for redesigning more systems. Therefore, based on utilizing the existing software, if we want to break the barriers and realize seamless integration using information technology, the inter-operation in heterogeneous environment must be solved, including the system maneuverability, normative description, communication protocol and security, etc.

#### 4) The problem of integrated mission planning

For space mission demands, mission planning needs to consider synthetically planning and resource scheduling of many systems. There are four situations of space mission demands: the first is that space mission with space-based TT&C terminal needs support of space-based TT&C system, ground-based TT&C system and payload TT&C system. The second case is that rocket with space-based TT&C terminal needs support of rocket TT&C system and space-based TT&C system. The third case is that military space mission needs joint support of rocket TT&C system, ground-based TT&C system, space-based TT&C system, payload TT&C system and space surveillance system for rapid combat response. The last case is that round-based TT&C system or space-based TT&C system supports emergency operation with the combination of space surveillance system. Therefore, there are a lot of problems to solve, such as the logical relationship between integrated mission planning and each system mission planning, comprehensive task decomposition method, model description, estimation algorithm, simulated demonstration, etc.

## 5 Suggestions for the Steps of Development

Development of the integrated concept of the ground segment for space internet includes three stages:

First stage: Determine space mission operation scene

Scene determination is to describe the requirements of future space mission operation from inter-connection, inter-communication and inter-operation based on mission target. Combined with the existed questions of China's space missions and the demand of future space missions, Academician Shen Rongjun proposed the overall objective, the composition, the network system architecture, and the network protocols of Integrated Space-Ground network system. He described the future space mission operation scene mainly from inter-connection and inter-communication, but inter-operation was not described. The premise of requirement description from inter-operation is to determine mission objective. For example, the objective of the U.S. Operationally Responsive Space (ORS) is to getting launch prepared in a few hours after receiving command and executing military campaign task in a few hours after getting to orbit. It also separates the overall objective into three parts. The objective could be the overall objective of China's space internet or the objective of space war in the future, and be the combination of the two objectives.

The second stage: break through key problems

From the aspect of united management, efficient sharing, cooperation and synthesis utilizing, the paper puts forward four key problems about space mission: architecture, information sharing and exchange, efficient cooperation cross system, and integrated mission planning. The key problems include many technologies, such as architecture research and optimization, network protocols of Integrated Space-Ground network system, network information management and security protection, space mission technology based on conceptual framework, integrated and complex mission planning of multi-system, etc. On the basis of independently simulating each key technology, the integrated ground simulation platform provides the validation platform for achievements.

The third stage: experiment on-orbit

In this stage, a new concept experiment of space mission is carried using on-orbit satellite, which is composed of three processes such as single system experiment, systems experiment, and multi-system joint experiment. Among them, the development and evolution of the conceptual framework needs to be considered in the design of the experiment systems, from the MOM+SOA to the complete SOA based on Coarse-grained subsystem, and to the complete SOA based on fine-brained component level, the shared services from Partly to Fully, and to cloud services.

**Acknowledgments.** The authors would like to thank the many participants of Intelligent mission planning and scheduling team in State Key Laboratory of Astronautic Dynamics. Part of the research described in this paper was sponsored by State Key Laboratory of Astronautic Dynamics.

Jian-ping Liu received the B.S., Ph.D. degrees from National University of Defense Technology, Changsha, China, in 1998 and 2005, respectively. His research interests include Intelligent mission planning and scheduling, Space Internet etc. Tel:+8602984762270. Email: ljpnudt@sina.com

## References

1. Bhasin, K., Hayden, J.L.: Space Internet Architectures and Technologies for NASA Enterprises. *International Journal of Satellite Communications* 20(5), 311–332 (2002)
2. Shen, R.: Some Thoughts of Chinese Integrated Space-Ground Network System. *Engineering Science* 8(10), 19–30 (2006)
3. Smith, D., Grubb, T., Esper, J.: Linking and Combining Distributed Operations Facilities Using NASA's "GMSEC" System Architecture. In: *SpaceOps 2008 Conference*, AIAA-2008-3483 (2008)
4. Sullivan, T., Sather, D., Nishinaga, R.: A Flexible Satellite Command and Control Framework. *Crosslink Summer*, 24–29 (2009)
5. Sarkarati, E., Spada, M., Moulin, S., Fischer, D.: Leveraging Advanced Software Technologies for Implementing the European Space Situational Awareness Ground Data Systems. In: *SpaceOps 2010 Conference*, AIAA-2010-1922 (2010)

# Related-Mode Attacks on PMAC

Jin Xu<sup>1,2</sup>, Qiaoyan Wen<sup>1</sup>, and Dayin Wang<sup>3</sup>

<sup>1</sup> State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup> Beijing Electronic Science and Technology Institute, Beijing 100070, China

<sup>3</sup> SKLOIS Lab, Institute of Software, Chinese Academy of Sciences, Beijing 100080, China  
xujinhappy@163.com, wqy@bupt.edu.cn, haowangyin@gmail.com

**Abstract.** PMAC(Parallelizable Message Authentication), proposed by Black and Rogaway in Eurocrypt'02, is fully parallelizable block-cipher mode of operation for authentication. Under the traditional attack models, PMAC is provable security. This paper uses related-mode attack to analyse the security of PMAC, and points out PMAC is insecurity when we use a few block cipher modes of operate at the same time using the same key, and gives some results and some advices on how to use it.

**Keywords:** Block cipher, Modes of operation, Related-cipher attack, Related-mode attack.

## 1 Introduction

The concept of Related-mode attack presented by W. Phan in [1] comes from the related-cipher attack presented by Wu[2].

**Related-cipher attack:** Block ciphers are often proposed with several variants, in terms of a different secret key size and corresponding number of rounds. Wu [2] presented the related-cipher attack model applicable to related ciphers in the sense that they are exactly identical to each other, differing only in the key size and most often also in the total number of rounds.

**Related-mode attack:** In [1], the authors generalize the concept of the related-cipher attack model to apply to a larger class of related model, in particular cipher encryptions with different block cipher modes of operation, but with the underlying block cipher being identical. They called it related mode attack and further show that when the adversary has access to an oracle for any one mode of operation of ECB, CBC, OFB, CFB, then almost all other related cipher modes can be easily attacked. It is reasonable to get an encryption and decryption oracle. For example, When A.Joux[3] attacked the EMD[4] mode of operation, he used an encryption and decryption oracle.

In this paper, we will discuss how to use any one of standard block cipher encryption modes to attack PMAC[5] under the related-mode attack model. In Section 1 and 2, we briefly describe the standard block cipher modes of operation and PMAC. In Section 3, we discuss how to attack PMAC. We conclude in Section 4.

## 2 Standard Block Cipher Encryption Modes

When encrypting a plaintext  $P$ , which is longer than the block size,  $n$  of the underlying block cipher, this plaintext is divided into  $m$  number of  $n$ -bit blocks  $P_i$ , and each one is encrypted at a time using a block cipher mode of operation that includes the Electronic Code Book (ECB), the Cipher Block Chaining (CBC), the Cipher FeedBack (CFB), the Output FeedBack (OFB) and the Counter Mode (CTR) [6,7].

The ECB mode is the simplest, where each plaintext block  $P_i$  is independently encrypted to a corresponding ciphertext block  $C_i$  via the underlying block cipher  $E_K$  keyed by secret key  $K$ :

$$C_i = E_K(P_i)$$

Figure 1 illustrates the ECB mode encryption on two consecutive plaintext blocks  $P_{i-1}$  and  $P_i$ .

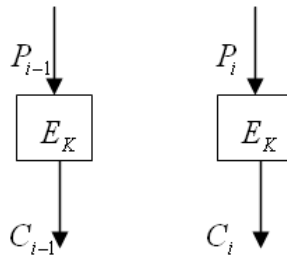


Fig. 1. ECB mode encryption

Meanwhile, the CBC mode uses the previous ciphertext block  $C_{i-1}$  as the feedback component that is eXclusive-ORed (XORed) to the current plaintext block  $P_i$ , before the resulting XOR is encrypted to obtain the current ciphertext block  $C_i$ . In particular:

$$C_i = E_K(P_i \oplus C_{i-1})$$

where  $C_0 =$  initialization vector (IV). Figure 2 illustrates the CBC mode encryption on two consecutive plaintext blocks  $P_{i-1}$  and  $P_i$ .

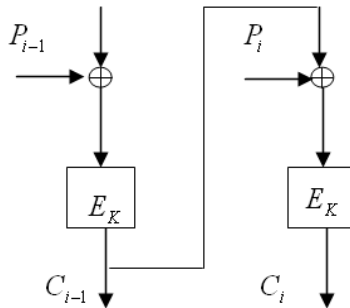


Fig. 2. CBC mode encryption



The CFB mode also uses the previous ciphertext block  $C_{i-1}$  as feedback, which is first encrypted and then XORed to the current plaintext block  $P_i$  to obtain the current ciphertext block  $C_i$ :

$$C_i = P_i \oplus E_K(C_{i-1})$$

where  $C_0 =$  initialisation vector ( $IV$ ). The CFB mode can also be viewed as a stream cipher mode by treating  $X_i = E_K(C_{i-1})$  as a keystream that is XORed to the plaintext  $P_i$  to obtain the ciphertext  $C_i$ . Figure 3 shows the CFB mode.

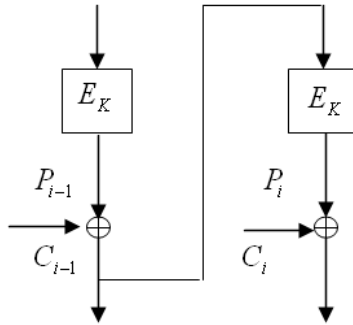


Fig. 3. CFB mode encryption

The OFB mode is similar to the CFB in that a keystream is also generated to be XORed to the current plaintext block  $P_i$  to obtain the current ciphertext block  $C_i$ . The difference is that the keystream is not a function of the previous ciphertext block  $C_{i-1}$ , but is the previously encrypted feedback component  $X_i$ :

$$X_i = E_K(X_{i-1}) \quad C_i = P_i \oplus E_K(X_i)$$

where  $X_0 =$  initialisation vector ( $IV$ ). Note that the keystream is independent of previous plaintext and ciphertext blocks. Figure 4 illustrates the OFB mode.

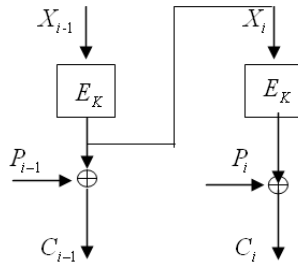


Fig. 4. OFB mode encryption

The CTR mode is similar to the CFB in that a keystream is also generated to be XORed to the current plaintext block  $P_i$  to obtain the current ciphertext block  $C_i$ . The difference is that the keystream is a function of a counter,  $ctr$ , which can also be looked on as an initialisation vector. Figure 5 illustrates the CTR mode.

$$C_i = P_i \oplus E_K(ctr + i)$$

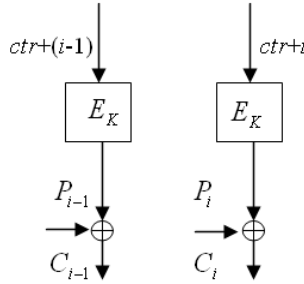


Fig. 5. CTR mode encryption

There are two variants of the mode, one random and the other stateful. No matter which variant is used, the initialisation vector,  $ctr$ , is included in the ciphertext as the first block  $C_0$  in order to enable decryption. The counter is not allowed to wrap around. Thus the decryption algorithm first chops off the first  $n$  bits  $C_0$  and uses it as  $ctr$ , and then divides the rest of the string into  $n$ -bit blocks and decrypt ciphertext using the same method of encryption.

In the following, we will discuss the definition of PMAC and how to attack PMAC under the related-mode attack model.

### 3 PMAC

PMAC is a Parallelizable Message Authentication Code. We say a Message Authentication Code like PMAC is security if any adversary can't find two different messages  $M$  and  $M'$  to make  $PMAC_K(M) = PMAC_K(M')$ .

Figure 6 illustrates the definition of PMAC[5].where  $E_k$  is a  $n$ -bit block cipher.  $L = E_k(0)$ ,  $\gamma_i$  is a gray code,  $x$  is a constant in  $GF(2^n)$ ,  $L \cdot x^{-1}$  is the multiplication operator in  $GF(2^n)$ .

According to the length of message  $M$ , there are two kinds of situation: if  $M$  is the integral multiple of  $n$ , i.e.  $M \in (\{0,1\}^n)^+$ , the last block should xor  $L \cdot x^{-1}$  before encrypted by  $E_k$ . If  $M$  is not the integral multiple of  $n$ , the last block of  $M$  should be padded to  $n$  by  $10^i$  ( $i = n - 1 - |M| \bmod n$ ), and then xor with  $0^n$ .

For PMAC, if we can get the value of  $L$ , then we can get  $L \cdot x^{-1}$ , so we can do the following forgery :

Let  $|M|=3n$ ,  $M = M[1]M[2]M[3]$ ,  $T = \text{PMAC}_K(M)$ , If we know the value of  $L$ , we can construct a new message  $M' = M[1]M[2]M'[3]$ ,  $|M'[3]| < n$ , and  $\text{PMAC}_K(M')$  is still  $T$ .

We can construct  $M'$  like this. Let  $M[3] \oplus L \cdot x^{-1} = N$ , then let  $M'[3]$  equals  $N$  truncate the last  $10^i$  string. After padding,  $M' = M[1]M[2]N$ ,  $M[3] \oplus L \cdot x^{-1} = N$ , since the first two block is the same, the input of the last is the same, then the output of  $\text{PMAC}_K(M')$  is still  $T$ , so  $M'$  is a new forgery.

From above we know if we can get the value of  $L$ , we can get a new forgery. In the following, we will try to get the value of  $L$  under the related-mode attack model, so we can use it to get a new forgery.

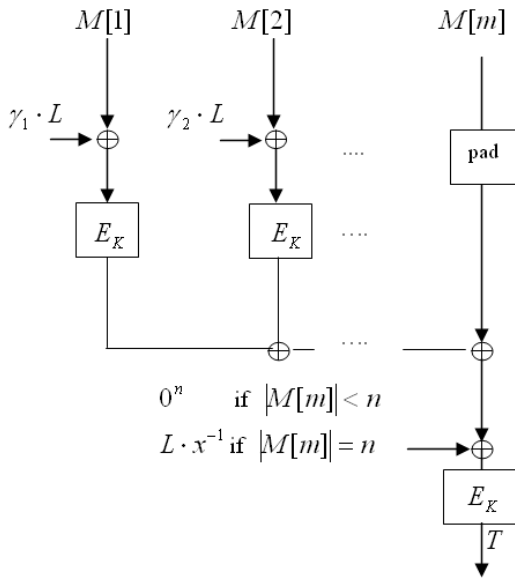


Fig. 6. PMAC Message Authentication Code

### 4 Using Standard Modes to Attack PMAC

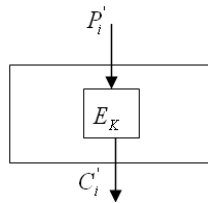
Throughout this paper, we consider the case where the adversary has access to an oracle that is able to perform either encryption or decryption for some fixed mode. This is similar to having access to known or chosen plaintext/ciphertext queries under that mode. We show that this oracle allows the adversary to attack other related-cipher modes, where the underlying block cipher is the same.  $P_i'$  and  $C_i'$  respectively denote the current plaintext and ciphertext block used in the interaction with the oracle being exploited,  $IV$  is the initialization vector. For the mode being exploited, access to its oracle allows the adversary to obtain known or chosen plaintext/ciphertext queries, and as necessary known or chosen IV queries - though

we assume for more concrete and interesting results that he can only access either a mode encryption or a mode decryption oracle, and not both at the same time. We will show how this oracle can be exploited to obtain a forgery under the PMAC mode.

**Exploiting an ECB Oracle**

Consider that the adversary has access to an encryption oracle under ECB mode, he only need to let  $P_i'=0$  and get the ciphertext  $C_i'$ , which is  $L$ . This is illustrated in Figure 7.

In summary, we require just one chosen plaintext(CP) query under ECB to obtain the  $L$ . So using the above method, we can get a new forgery.



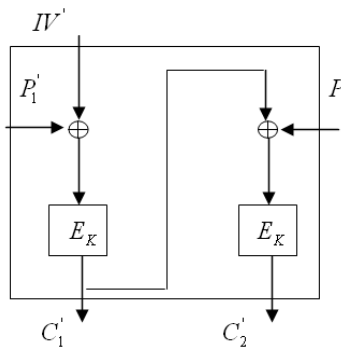
**Fig. 7.** Exploiting ECB to get  $L$

**Exploiting a CBC Oracle**

When the adversary has access to a CBC oracle, he only need one chosen plaintext query under known  $IV$  defined in [8] to attack PMAC.

The adversary let  $P_1' = IV'$ , and query  $P_1'$ , then the output of CBC is  $L$ , because  $C_1' = E_K(P_1' \oplus IV') = E_K(0) = L$ . This is illustrated in Figure 8.

In summary, we require just one chosen plaintext(CP) query under CBC to obtain  $L$ .



**Fig. 8.** Exploiting CBC to get  $L$

### Exploiting a CFB Oracle

The adversary accesses a CFB decryption oracle and chooses  $C'_{i-1} = 0$ , and hence  $E_K(C'_{i-1}) = E_K(0)$ . since  $E_K(0) \oplus C'_i = P'_i$ , then  $E_K(0) = P'_i \oplus C'_i$ . See Figure 9.

In summary, we require just one chosen ciphertext (CC) query under CFB to obtain  $L$ .

### Exploiting an OFB Oracle

This is so far the hardest attack to mount, and requires a chosen-IV (CIV) scenario [9]. In particular, the adversary chooses  $IV' = 0$ , and hence  $E_K(IV') = E_K(0)$ ,  $E_K(0) = P'_1 \oplus C'_1$ . Then he get  $L$ . See figure 10.

In summary, we require just one chosen ciphertext (CC) query and one chosen-IV query under CFB to obtain  $L$ .

Note that in this case the plaintext and ciphertext blocks of the exploited oracle do not need to be chosen but are merely known.

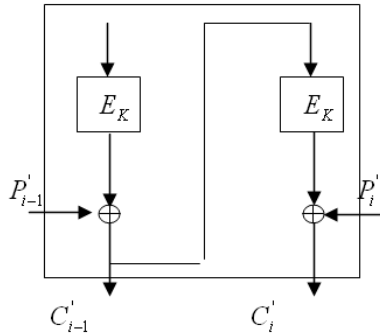


Fig. 9. Exploiting CFB to get L

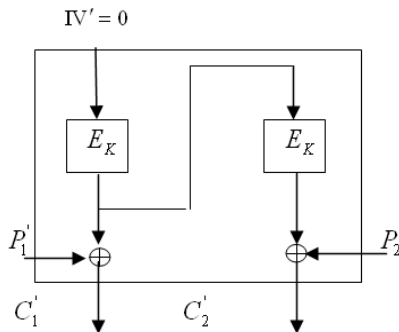


Fig. 10. Exploiting OFB to get L

### Exploiting a CTR Oracle

In this situation, the adversary need to access CTR decryption Oracle. For any cipertext  $C$ , Let  $C'_0 = -1$ , since  $P'_1 = E_K(C'_0 + 1) \oplus C'_1 = E_K(0) \oplus C'_1$ , then  $E_K(0) = C'_1 \oplus P'_1$ . See figure 11.

In summary, we require just one chosen ciphertext (CC) query under CTR to obtain  $L$ .

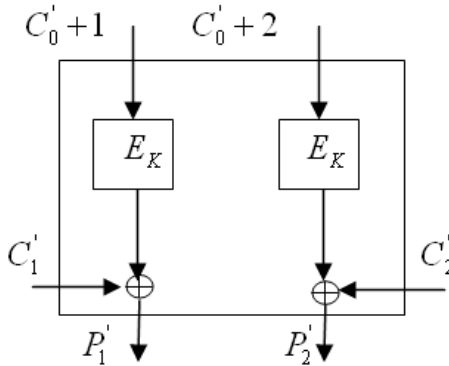


Fig. 11. Exploiting CTR to get L

## 5 Conclusions

In this paper we discuss how access to chosen plaintexts/ciphertexts Oracle in standard encryption modes allows related-cipher PMAC mode to be attacked. In Table 1, we list our attacks and the corresponding text complexities, while computational complexity is negligible.

Table 1. The complexities of Related-mode attacks on PMAC

Oracle Exploited	Cipher Mode Attacked	Text Complexity	Using Oracle
ECB	PMAC	1CP	ECB Encrypt
CBC	PMAC	1CP	CBC Encrypt
CFB	PMAC	1CC	CFB Decrypt
OFB	PMAC	1CP,1CIV	OFB Encrypt
CTR	PMAC	1CC	CTR Decrypt

1CP means one chosen plaintext query; 1CC means one chosen ciphertext query;1CV means one chosen IV query;

There are a lot of standard authentication modes and authenticated encryption modes using block ciphers, such as authentication mode CMAC[10], authenticated

encryption mode GCM[11]. We further study the security of those modes of operation under related-mode attack model and find they all are insecure if the adversary can access to an oracle under one proper mode. So when we have the same cipher being used as the under-lying component in different block cipher modes of operation, we should avoid using the same key in those modes in practical applications.

**Acknowledgments.** This work is supported by NSFC (Grant Nos. 60873191, 60903152, 61003286, 60821001), the Fundamental Research Funds for the Central Universities (Grant Nos. BUPT2011YB01, BUPT2011RC0505).

## References

1. Phan, R.C.W., Siddiqi, M.U.: Related-Mode Attacks on Block Cipher Modes of Operation. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3482, pp. 661–671. Springer, Heidelberg (2005)
2. Wu, H.: Related-Cipher Attacks. In: Deng, R.H., Qing, S., Bao, F., Zhou, J. (eds.) ICICS 2002. LNCS, vol. 2513, pp. 447–455. Springer, Heidelberg (2002)
3. Joux, A.: Cryptanalysis of the EMD Mode of Operation. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 1–16. Springer, Heidelberg (2003)
4. Rogaway, P.: The EMD Mode of Operation. Cryptology ePrint archive[EB/OL] (2002), <http://eprint.iacr.org/2002/148/>
5. Black, J., Rogaway, P.: A Block-Cipher Mode of Operation for Parallelizable Message Authentication. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 384–401. Springer, Heidelberg (2002)
6. FIPS 81, National Institute of Standards and Technology. DES Modes of Operation (1980)
7. SP 800-38A, National Institute of Standards and Technology. Recommendation for Block Cipher Modes of Operation – Methods and Techniques (2001)
8. Hong, D., Sung, J., Hong, S., Lee, W., Lee, S., Lim, J., Yi, O.: Known-IV attacks on triple modes of operation of block ciphers. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 208–221. Springer, Heidelberg (2001)
9. Wagner, D.: Cryptanalysis of Some Recently-proposed Multiple Modes of Operation. In: Vaudenay, S. (ed.) FSE 1998. LNCS, vol. 1372, pp. 254–269. Springer, Heidelberg (1998)
10. National Institute of Standards and Technology (NIST), NIST Special Publication 800-38B, Recommendation for Block Cipher Modes of Operation: The CMAC Mode for Authentication [EB/OL] (May 2005)
11. McGrew, D., Viega, J.: The Galois/Counter Mode of Operation (GCM), Submission to NIST Modes of Operation Process [EB/OL] (2004), <http://csrc.nist.gov/CryptoToolkit/modes/proposedmodes/>

# Process Management of Civil Aircraft Development Project Based on Workflow

Zeng Yuqin, Yan Guangrong, and Zhao Gang

D316 New Main Building, 37# XueYuan Rd, HaiDian Dist, Beijing, China  
zengyuqin530@yahoo.com.cn, yangr@buaa.edu.cn, zhaog@buaa.edu.cn

**Abstract.** Beginning with the problems emerged in project implementation, the paper proposes a project process management solution based on workflow after analyzing the features and demands of process management in aircraft development projects. We use hierarchical network plan diagram to picture the running process of the project and put forward a kind of process model, process driving and monitoring techniques for civil aircraft project. Besides, we design a project management system based on the research and successfully apply it to i-Plane platform.

**Keywords:** civil aircraft development project, process management, workflow, process driving, process monitoring.

## 1 Introduction

After having completed project planning, it's necessary to ensure that activities in the plan run with the setting time and logic. And the implementation process should be monitored. These are the main content of process management research. [1] Nowadays, the aviation project management system usually uses some artificial methods to push the project planning. Also we monitor the running process of the project just only by setting some milestone tasks in the critical path of a plan. However, since civil aircraft development projects have characteristics of large scale product data, complex organization structures, and numerous operational tasks to be performed during a long development cycle, this method can't reflect the implementation effect of the project accurately and is lack of project monitoring and timely controlling. It will seriously affect the running efficiency of the civil aircraft project. [2] Meanwhile, the workflow driven engine can provide a technical solution to promote the project activities. In this paper, we propose a process model on the basis of workflow technology for civil aircraft project, and use workflow technology to drive and monitor the project process.

## 2 Project and Process Management

### 2.1 Needs of Process Modeling

The process model for civil aircraft development project must meet its requirements. This kind of projects has the following characteristics [3]:



① The planning of the project establishment and implementation. The establishment and implementation of civil aircraft project are planned very strictly. It is Often based on the market demand, and issued as a clear mandate by the higher authorities, with strong planning;

② The hierarchical structure of the management organization. The management organization of civil aircraft project is usually a three-level structure which includes group companies, research institutes, and plants. This also determines the project plan to be a hierarchical network;

③ The distribution and collaborative of the participating research units. The high complexity and multi-disciplinary of civil aircraft projects determine the breadth of the participating research units. It needs multiple units and staffs which are in different areas or different departments work together. So how to ensure the smooth flow of information, project implementation and timely feedback to smooth the collaboration must be considered as building process model.

### 2.2 Hierarchical Network

Due to the hierarchical structures of aircraft products and management organization, the project plan is always hierarchical. [4] As the total project plan is broken down to sub-plans layer by layer, the hierarchical network plan can be formed.

To formulate the multi-layer network plan, firstly, we should divide the level of the network plan. We use the method which is commonly used in aviation project and decomposes tasks top-down and layer by layer to divide the network plan into four layers. Figure 1 uses a hierarchical network diagram to show project plan pictured with a tree structure.

- ① Layer 0: The project primary node and milestones generally identified by the Group Company after determining the project’s objective;
- ② Layer 1: Departmental implementation plans specifically developed by related departments / research units based on tasks they have receipted;
- ③ Layer 2: Program plan made for key or important parts of the plane.
- ④ Layer 3: The manufacturing and assembly sequence of the products which are detailed to each process of a part and a component as well as the corresponding person in charge.

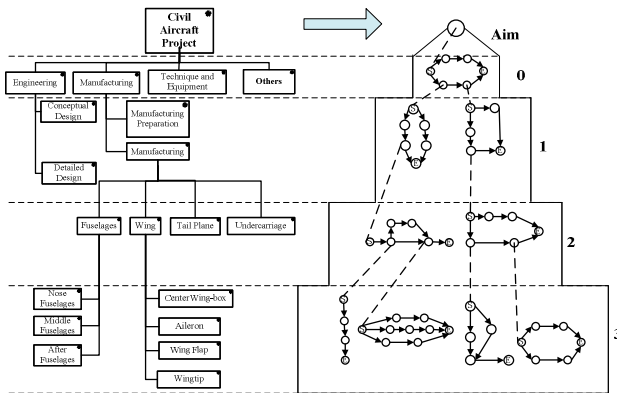


Fig. 1. WBS to Hierarchical Network Plan

When we establish a multi-layer network planning in practice, the project can be further subdivided on the basis of the third layer until the task can be broken down into more specific activities to the duty person. After being completed the top-down task decomposition, the network plan must be reviewed from bottom to up. Only the multi-layer network plans which have passed the examination can be started.

### 3 Workflow Model

#### 3.1 The Function and Structure of Workflow Engine

[5] According to the definition of Workflow Management Coalition, standard workflow model includes workflow engine and five interfaces (process defining tools, client application tools, calling program, other workflow implementation services, and managerial monitoring tools). However, as for project process management, only process driving and monitoring are necessary. Therefore, under the demands of project process management, a structural framework of the workflow showed in Figure 2 is put forward after simplifying some interfaces. Functions of each part are defined as follows:

① User interface: User interface is based on Web Browser to accomplish the interaction with the system;

② Planning manager: Planning manager is responsible to define various information of project process;

③ Process manager: Process manager is responsible to manage and create process example, to define the relationship and the relative parameter (time parameter, person in charge and resources) among the Web activities and to form activity examples which are understandable to process manager;

④ Process monitor: Process monitor is to have access to the beginning and ending time of the activities and the relationship among them, to judge whether the activity beginning or ending is normal, to send feedbacks to drive engine and to urge the process rate when it is not in agreement with plan;

⑤ Drive engine: Drive engine is to adjust the status of activity or process example based on the schedule of activity or process, user input and the monitored information by process monitor. It is responsible for the state changes, for driving the normal operation according to the timing of network nodes and for the feedback of driving condition towards process monitor;

⑥ Resource manager: Resource manager is to provide the resources to manage the projects for project operation and to put the newly added resources into the repository;

File manager: File manager is to manage the output files generated by various activities, to put them into the project file repository and to provide corresponding input file for drive engine when needed.

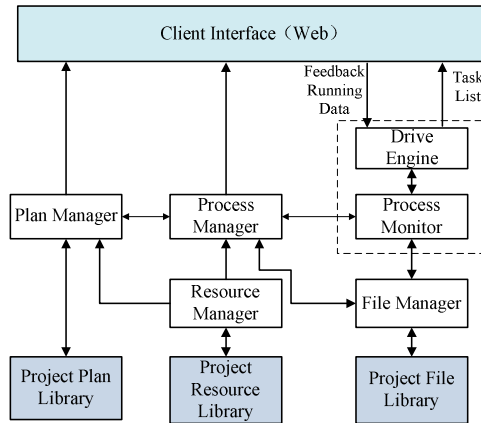


Fig. 2. Structural framework of the workflow

### 3.2 The Mechanisms of Workflow Model Driving and Monitoring

The driving and monitoring mechanism based on workflow is essentially a workflow server coring the workflow engine [5]. It is interactive with other application interfaces so that activities are implemented in the order set by the network planning. It is to accomplish the activity state changes and date collection as well as the access to external resources as required.

## 4 Process Driving and Monitoring

The above workflow structural framework and driving mechanism serve the implementation and driving of project process as foundation. The operation and driving of process instantiations are discussed as follows.

### 4.1 Instantiations of Workflow Process

In hierarchic network model, digraph is used to express the operation process of project instantiations. Nodes are activity instantiations in the project and directed edges are the relationship among activities.

**Definition 1.** Time constraints.  $T$  is defined by a five- element equation:  $T = \{ designStartT, designEndT, duration, actualStartT, actualEndT \}$ . It is explained in Table 1.

**Table 1.** Instructions of *T*

<b>Name</b>	<b>Descriptions</b>
<i>T</i>	time constraints
<i>designStartT</i>	the estimated start time of <i>activityE</i>
<i>designEndT</i>	the estimated end time of <i>activityE</i>
<i>actualStartT</i>	the actual start time of <i>activityE</i>
<i>actualEndT</i>	the actual end time of <i>activityE</i>
<i>duration</i>	the delay of <i>activityE</i> $duration = [(designEndT - designStartT) - (actualEndT - actualStartT)]$

When implementation is delayed (or  $duration < 0$ ), activity duty person will receive a message or warning.

Implementation detention will directly influence the relative following-up timing, especially for hierarchic model of network planning. During a project operation process, detention is inevitable. There must be relevant response measures: making use of time service to grasp and urge the duty person, simplifying the flow and cancelling the unnecessary alternative activities when drafting the hierarchic network planning.

**Definition 2.** *Status* is to represent the status of *activityE* or *processE*. There are five kinds of values: *ready*, *running*, *suspended*, *waiting* and *completed*.

**Definition 3.** Activity instantiation (*activityE*) is component of the flow and in the hierarchic model of network planning is a node of process instantiation (*processE*) which can be defined as a seven-element equation:  $activityE = \{nameA, T, Resource, finishPercent, dutyperson, status, File\}$ . It is illustrated in Table 2.

**Table 2.** Instructions of *activityE*

<b>Name</b>	<b>Descriptions</b>
<i>T</i>	time constraints
<i>nameA</i>	the name of <i>activityE</i>
<i>Resource</i>	the instantiation of resource be used by <i>activityE</i>
<i>finishPercent</i>	the completed percent of <i>activityE</i>
<i>dutyperson</i>	staffs or departments in charge of <i>activityE</i>
<i>status</i>	the status of <i>activityE</i>
<i>File</i>	files related to <i>activityE</i> $File = \{inFile, outFile\}$

**Definition 4.** Process instantiation (*processE*) is one implementation process of the project which means one bottom network operation of the hierarchic network

planning. It is defined as a five-element equation:  $processE = \{namePr, nameP, T, A, status\}$ . It is demonstrated in Table 3.

**Table 3.** Instructions of  $processE$

Name	Descriptions
$namePr$	the name of $processE$
$nameP$	the name of Project, $processE$ belongs to it
$T$	time constraints
$A$	the set of all activity instantiations which are generated during $processE$ running
$status$	the status of $processE$

**Definition 5.** Task ( $taskE$ ) is defined as the tasks which a certain user or organization is supposed to accomplish. It is represented by a three-element equation :  $taskE = \{processE, activityE, dutyPerson, finishValue\}$ .

$ProcessE$  is process instantiation;  $activityE$  is activity instantiation;  $dutyPerson$  is the responsible person;  $finishValue$  is the completed percent of the task. As the finished task percentage ( $finishValue$ ) is monitored and served as feedback, the activity progress rate is updated and responded to the upper level so that a bottom-up feedback mechanism is form in the hierarchic network model.

### 4.2 Process Driving and Monitoring

[6] The initial relationship among activities in project operation primarily include serial, parallel and coupling. The serial and the parallel can be defined in network planning while the coupling has to go through decoupling or merging and get transferred into the serial or the parallel to be defined. The following algorithms only focus on the serial and parallel activity relationship.

**Algorithm 1.** Create  $processE$ :

- ① Select appropriate network planning model  $P$  from network planning model repository under the demands;
- ② Create  $processE$  according  $P$  with initial parameter:  $namePr, nameP, T$ . The creation is finished.

**Algorithm 2.** Start  $processE$

- ① Judge whether the value of  $status$  is *ready*. If not, start is failed and the algorithm is finished; if true, enter ②);
- ② Assignment :  $status=running$ ;
- ③ Obtain the beginning activity of  $processE$  and create  $activityE$ ;
- ④ Implement  $activityE$ .

**Algorithm 3.** Implement activity example ( $activityE$ )

- ① Judge whether the activity node type is the origin sub-node  $a$ . If true, enter ②); if not, the activity node type is sub-process  $p$ . Read the sub-process’s status and obtain its starting activity node and then judge again until entering ② ;

- ② Judge whether all of the predecessor activity are completed ( $status==completed$ ) and input files ( $inFile$ ) are matched. If true, enter ③; if not, set its status to be waiting until being true;
- ③ Create a corresponding  $taskE$  for  $activityE$ . Judge whether  $taskE$  is implemented automatically. If true, apply the relative implementing function; if not, inform its  $dutyPerson$  to implement;
- ④ Obtain the  $finishValue$  of  $taskE$  from work item form and combine with the  $finishPecent$  of  $activityE$  to get the current value:  $finishPercent=finishPercent+finishValue$ ;
- ⑤ Judge whether the activity is finished. If  $finishPercent=100$ , the activity is finished, enter ⑦; if not, enter ⑥;
- ⑥ Judge whether the time constraints are fulfilled. If ( $designEndT<actualEndT$ ) or ( $finishPercent<100$ ), the time constraints are not fulfilled, send message to urge  $dutyPerson$  and record the current  $finishPercent$  and  $duration$ ; or, the time constraints are fulfilled, the algorithm is finished;
- ⑦ Judge whether the  $outFile$  and the conditions required by the rear activities are matched. If true, sent finish information to upper network node and obtain set  $A$  including the  $activityE$ 's rear activities and implement activity from ① according to the setting time; if not, send message to  $dutyPerson$ .

## 5 Implementation and Application

According to research on process driving and monitoring in civil aircraft project and workflow, the passage aims to discuss how to implement aviation research project management system. Figure 3 shows the specific project management process.

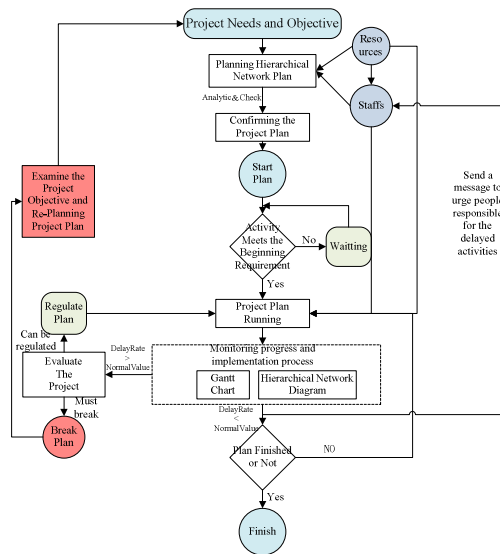


Fig. 3. Project management process

The project management system discussed in the paper is acted as a sub-system applied to i-Plan platform. I-Plan is a distributed and networked collaborative development platform for large aircraft. The project management system is mainly responsible for definition of program planning and implementation of process management. Figure 4 explains the role of the system performed in i-Plan platform.

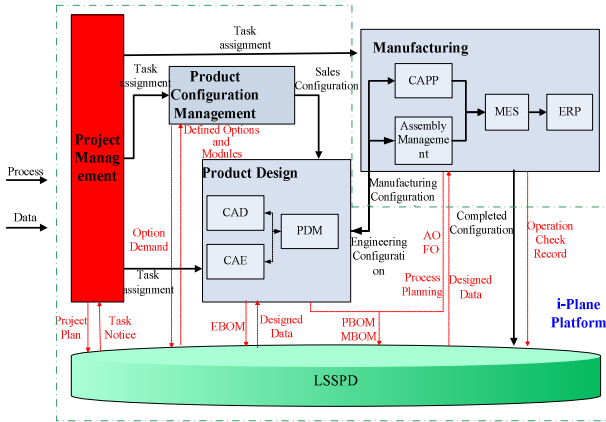


Fig. 4. Business process of i-Plane

## 6 Conclusions

In this paper we have presented a solution based on workflow technology to solve problems existing in the aircraft project implementation. In order to meet this kind of project’s demands and characteristics, we describe project plans by multilayer network plan diagram. Then a workflow process model was constructed after studying workflow driving mechanism, which can be used to establish a project management system. The system can help staffs realize process driving and monitoring, and has been applied in i-Plane platform to better manage aircraft development projects.

**Acknowledgments.** The author thanks China’s 863 program--“Key Technologies of i-Plane for Large Aircraft Development” for the sponsorship of this paper. The project number is 2009AA043306.

## References

1. Shen, J.-M., Bai, S.-J.: Introduction defense high-tech project management. Machinery Industry Press (2004)
2. Xu, Z.-Y., Zhang, K.-F., Li, Z.-L., et al.: Method of hierarchical network plan for complex aviation product. Computer Integrated Manufacturing Systems 12(005), 727–730 (2006)

3. Wang, C.: Research on Techniques of Multi-Project Planning and Controlling for Aviation Manufacturing. Northwestern Polytechnical University (2007)
4. Deelman, E., Gannon, D., Shields, M., Taylor, I.: Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems* 25, 528–540 (2009)
5. Torres, V., Pelechano, V.: Building business process driven web applications. *Business Process Management*, 322–337 (2006)
6. Juan, Y.C., Lin, J.S., Ou-Yang, C.: A Process-Driven Approach to Develop Multi-agent System for Cooperation in Concurrent New Product Development. *IEEE* (2009)



# JMF-Based Video Surveillance and Playback

Gao Xiuhui<sup>1</sup>, Wang Wei<sup>1</sup>, and Zhao Gang<sup>2</sup>

<sup>1</sup> D316 New Main Building, 37# XueYuan Rd, HaiDian Dist, Beijing, China

<sup>2</sup> B434 New Main Building, 37# XueYuan Rd, HaiDian Dist, Beijing, China  
Gaozi1012@126.com, jttr@buaa.edu.cn, Zhaog@buaa.edu.cn

**Abstract.** In order to solve the complex installation and low security in video surveillance system, based on the Java Media Framework (JMF) Application Programming Interface and the real-time transport protocol RTP, a B/S mode video surveillance system under the internet environment is presented. The system framework, implement process and the solution to the key problem are also introduced in details. The system is not only highly safe and easy to use, but also meets the requirements of cooperation and surveillance in plants or R&D department within the development process of aircrafts.

**Keywords:** video surveillance, JMF, real-time transport protocol.

## 1 Introduction

In recent years, with the great progress of the computer technology and the rapid development of the Internet, the network video transmission technology has been widely used. Based on this technology, a range of research areas, such as the video surveillance, video conferences, the video chat and the distance education, have widely developed, and increasingly played an important role.

In order to solve the key problems of real-time video transmission more conveniently, most of the current video transmission solutions of multi-media communication systems, have employed the MS Windows platform, which use Microsoft's DirectShow technology development or the MFC class library. Then, you can achieve better video transmission, and an ideal cost-effective. However, the application environment of these communication systems depends on the operating system and has poor portability. It cannot be cross-platform used and the installation process is relatively complicated. With the Java language platform-independent, secure, stable, multi-threading and other excellent features, Java based JMF(Java Media Framework) can be used not only to develop large applications, but also to develop the Internet applications [1]. This paper studies the JMF and RTP video transmission technology framework agreement, combined with specific applications to design and implement a JMF-based B/S mode video surveillance system. It has the video surveillance, access authentication and video playback features that allow users use it through Internet Browser directly.

## 2 Introduction of JMF Technology

JMF is introduced by the SUN Company. It can blind audio, video and other time-based media data to Java Application and Java Applet in the application program interface (API). JMF defines a uniform architecture, messaging protocols and programming interfaces for media player, media access and media playback [2]. It supports for media capture, playback, network streaming, and multiple media format conversion. At the same time JMF provides a programming abstraction mechanism for Java in the media to hide implementation details of the developer. JMF includes two parts, which are JMF API and RTP API. The main function of the former part is to capture, process, store and play media, and the latter one is mainly to transmit and receive media streams in the network [3].

Real-time Transport Protocol (RTP) is a transport protocol for Internet multimedia data streams. RTP applies to the transmission of one or many, the purpose of the situation is to provide multimedia data streams' synchronization, RTP typically uses UDP to transfer data, but it also works on TCP, ATM or other protocols [4]. When the application starts an RTP session, it will use two ports. One port is assigned to the RTP, and the other one is assigned to RTCP. RTP does not provide any mechanism to ensure the transmission quality and does not provide flow control or congestion control. Services of transmission quality and reliability rely on RTCP protocol.

JMF-based multimedia data streams of RTP transmission, reception and playback process are shown in Figure 1 [5]. Data can be collected from multimedia files, such as video files. It can also be collected from the audio or video equipment, such as audio grabber and multi-media cameras. These data sources are in different positions, different formats, JMF locate the corresponding data source by a called class of media locator (MediaLocator). MediaLocator objects store the data of source location information. JMF manager (Manager) creates data source based on the information provided by the media locator. The data source is different from the physical data source of files, cameras. It is an abstract concept, that an abstraction of a variety of different types of data source. Thus, JMF can hide the underlying details, and developers can no longer consider the specific source and location data in the programming process, only consider the data format, speed and other information.

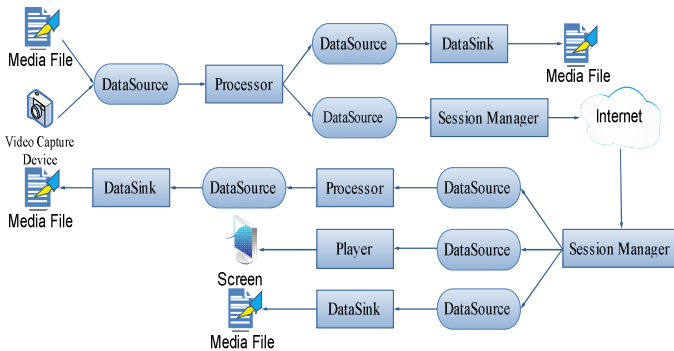


Fig. 1. JMF-based multi-media stream transmission process

### 3 Design of Video Surveillance System

#### 3.1 System Architecture

The surveillance system is based on Browser/Server (B/S) mode. Monitor front-end act as the role of server, all application software development, and maintenance work are focused on the server side, server side is equipped with audio/video capture hardware devices such as audio grabber and multi-media cameras etc. The functional architecture of surveillance system is shown in Figure 2. Server side includes five functional modules: the video transmission module, responsible for collecting multi-media data stream and packaged into RTP stream that can transmit on the network; the video storage module, responsible to collect the multi-media data stream to save into the local computer for the needs when the user uses playback; the video query module, responsible for listing the video files saved on the server side when the user uses playback; the permission management module, responsible for verifying the user's identity information to ensure the security of the system; the surveillance module, responsible for surveillance the client's request, such as video surveillance or video playback, surveillance each request corresponds to the different transaction processing thread. Client side has two main modules: the video surveillance module and the video playback module. The former module provides the video information to the user, and the latter one provides users the video information which is saved on the server side.

The client side, with the installation of the necessary protocols software, browser and connection to the network, can use the video surveillance system. And the user can select the video files previously saved on the server side to playback.

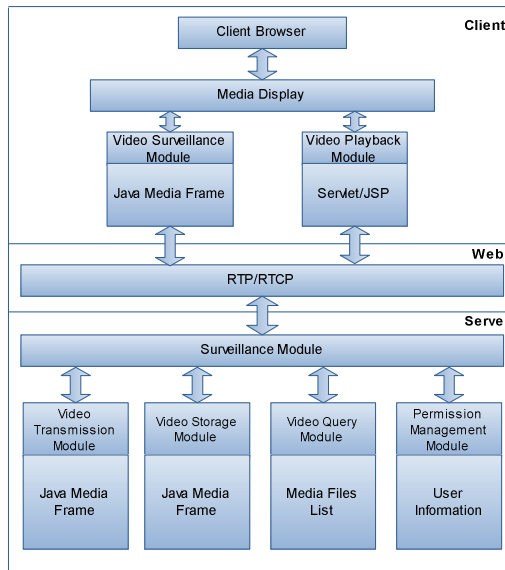


Fig. 2. Functional architecture of surveillance system

### 3.2 System Working Process

After booting the server computer, video storage module will run automatically, and in order to ensure the accuracy of searching files when playback, set every five hours the video auto-save as a video file. When client side needs video surveillance, the client application establishes connection with the server first, then can launch requests to server. Server side verifies the user's information through permission management module first, and builds UDP-based connection with the client, then determine the type of user's requests, if the request is video surveillance, put the collected multi-media data stream to specify the data format compressed by RTP encapsulation, and then send the multi-media data streams via RTP session to the client. The client will receive and decompress the multi-media data stream to play with the media display module; if video playback request, the server will send the video file lists to client through video query module, when the user chooses one of the video files, server will create multi-media data stream based on this file by JMF and then transmit it to the client, at last the client calls the video display module for video playback. The video working process is shown in Figure 3.

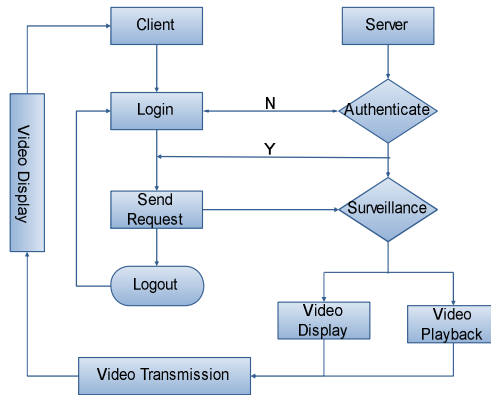


Fig. 3. System working process

## 4 Key Technology of System Implement

### 4.1 Capture and Compression of Multi-media Data

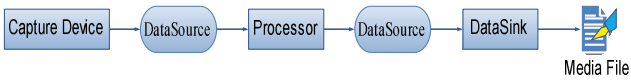
Video data is captured by video capture devices, and transmitted through RTP protocol. With the traditional AVI or MOV format, the server's pressure will increase and network requirements are particularly high, so it is necessary to transfer the traditional format into a good encoding which is easy to transmit, and has good performance on network adaptability and against packet loss performance. The multi-media compress formats supported by JMF when transmitted with RTP protocol are listed in table 1 [6]. Multi-media data stream has to make a conversion for network transmission after the output stream is tailored.

**Table 1.** Media transmit format supported by JMF

Media Type	Transmission format with RTP
Audio	AUDIO_G711_ULAW/RTP, DVI/RTP, G723/RTP, GSM/RTP
Video	JPEG/RTP, H.261/RTP, H.263/RTP

## 4.2 Local Storage of Multi-media Data

In order to preserve the audio and video information on the server side and provide playback functionality, it is necessary to store video files on the server side. The DataSink, which is supported by JMF, can read data from the output data source of processor object and save as a local file. First we should create a MediaLocator object used to specify the location of the file you want to save, and then duplicate a data source from data capture devices and use it to create a processor, at last use the output data source of this processor to create DataSink and save in local files with the specify format of MediaLocator. The multi-media data storage process is shown in Figure 4.

**Fig. 4.** Multi-media data storage process

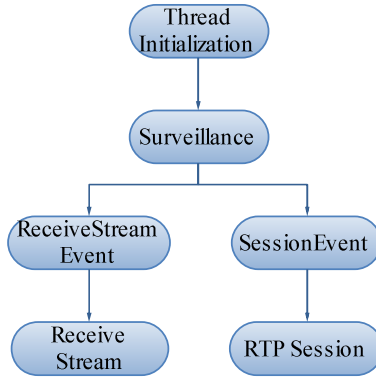
## 4.3 Reception and Play of Multi-media Data

The processing on client side includes reception, decompression and playback of multi-media data. When the client makes successful connection with server, it will begin to prepare to receive and play multi-media data stream. Receiving multi-media data stream also needs to create RTP session, and the process is similar with the process of sending data stream, but the difference is for each RTP manager it's necessary to add the listeners to listen for the situation of corresponding address and port, which include the RTP session and state of the RTP stream, RTP events can be divided into five categories:

- ReceiveStreamEvent
- RemoteEvent
- SendStreamEvent
- SessionEvent

ReceiveStreamListener listener monitors whether RTP media stream arrive, the ControllerListener listener monitors player' events, and SessionListener listener monitors session events, the receive process of multi-media data is shown in Figure 5.

When a listener monitors the data stream has arrival, and then create a player, while registering ControllerListener listener in the player to control the video playback on the player.



**Fig. 5.** Multi-media data receive process

#### 4.4 Playback of Multi-media Files

When the client side makes a successful connection with the server side, the client can choose to list the video files stored on the server, and select one of the files to playback.

The client side can send request to web server, which is on the server side, through Servlet technology, and then web server will return a list of video files at the local. When the client selects a file, the server will establish multi-media data stream based on the file the client chooses, and transmit the stream to client side through RTP protocol. The server obtains a URL by the filename client chooses, which is used to establish a MediaLocator includes the client's IP address and receive port for transmitting on network. With the MediaLocator, the client will receive multi-media data stream and play it instead of saving the files on client side first.

## 5 Example with Surveillance System

This study primarily relies on the prototype system named i-Plane (Information based Platform for Large Aircraft Network Environment), whose target is to provide a distributed collaborative platform for large aircraft development, and achieve personnel, data and processes in close collaboration. The hardware deployment framework of i-Plane is shown in Figure 6.

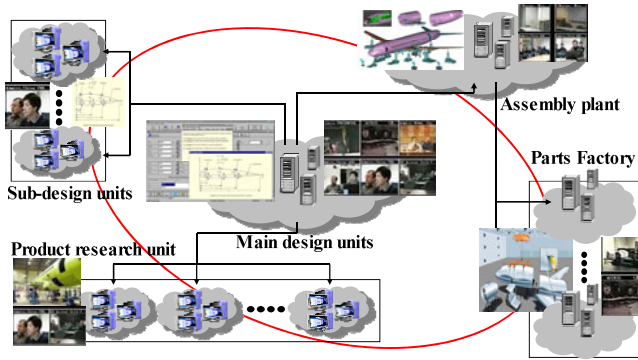


Fig. 6. Hardware deployment framework of i-Plane

In the collaborative process, the different units are in different locations. In order to safeguard the smooth progress of collaborative projects, the inter-units need use the surveillance system to monitor the project site. Figure 7 shows the program screenshots when use the surveillance system.

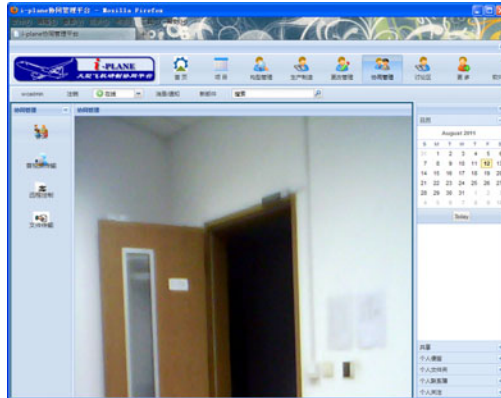


Fig. 7. Program screenshots of surveillance system

## 6 Conclusion

This article introduces the system architecture, work process and key technology of the JMF-based B/S mode surveillance system. The B/S mode surveillance system focuses on the combination of web technology and surveillance system, which simplify the installation of client software. Just with the installation of protocol software and connection with internet, the client can monitor within the permission of security at anytime and anywhere. This system can meet the requirements of surveillance in plants and cooperation in R&D departments within the development

process of aircrafts. Through the research to JMF technology, we found that the JMF technology has huge advantage in the settlement of multi-media data transmission. As the improvement of Java and JMF technology, the JMF technology will play an increasingly important role in multi-media transmission field.

**Acknowledgments.** The author thanks China's 863 program—"Key Technologies of i-Plane for Large Aircraft Development" for the sponsorship of this paper. The project number is 2009AA043306.

## References

1. Tian, J.-B.: BIE Hong-xia. Real-time Audio & Video Communication Based on JavaMedia Framework (2006)
2. Sun, Y.-L., Peng, B.: Java Network Programming Examples, pp. 219–277. Tsinghua University Press, Beijing (2003)
3. Zhu, J.-J.: Research and implementation of audio-visual conference system based on JMF 12 (2009)
4. Lo Iacono, L., Ruland, C.: Confidential Multimedia Communication in IP Network. Communication Systems, 516–523 (2002)
5. Chen, L.-D., Chen, W.-X.: Research of Real-Time Video Multicast System based on JMF. Microcomputer Information (December 2009)
6. <http://baike.baidu.com/view/209561.htm> [DB/OL]



# New Aspects of Phoneme Synthesis Based on Chaotic Modeling

Marius Crisan

Department of Computer and Software Engineering, “Politehnica” University of Timisoara,  
Romania  
marius.crisan@cs.upt.ro

**Abstract.** The paper discusses the possibility of phonemes synthesis starting from a dynamic model based on chaotic modulation of the phoneme’s elementals. The natural phonemes can be analyzed in terms of elemental patterns that repeat themselves with slight chaotic variations along the signal length. The elementals are modeled by a classical harmonic modulation process having the parameters influenced by chaotic signals. The experimental results proved rather satisfactory and support further research in this direction.

**Keywords:** phoneme synthesis, speech processing, time series analysis, chaotic dynamical systems.

## 1 Introduction

Speech synthesis continues to be an important research domain as long as the results obtained are expected to overcome the challenge of sound naturalness as uttered by human voice. Traditionally, the most common speech synthesis approaches are based on wave form concatenation, speech production model, and vocal tract simulation [1-3]. The concatenation method uses segments of human recorded speech that are strung together. The key factor in this approach is the selection of the optimal unit length. Phonemes and diphones are the very commonly used units in synthesis. Speech production modeling, on the other hand, is not based on any human speech sample, but uses acoustic models. These models are mainly based on a linear filter that simulates the vocal tract. The formant synthesizer is the main technique of this category [4]. The vocal tract transfer function is modeled by the simulation of formant frequencies and amplitudes. A waveform of artificial speech is created by using rules starting from some target speech parameters for each phoneme. The characteristic of this approach consists of producing a waveform as close as possible to the natural speech signal without the need to simulate any aspect of the human vocal tract. In contrast, the vocal tract simulation approach employs a physical or articulatory model of the speech production apparatus that reconstitutes the physical phenomenon of air flow through the phonatory organs. The classical model in this category is the two-mass representation of the vocal cords [5]. Although these linear techniques were refined in time and may offer a limited technological success in some applications, they proved inadequate for the most speech processing demands. The most striking limitation is the

production of very mechanical sounds which lack human emotion. On the other hand, there is strong experimental and theoretical evidence for the existence of nonlinear phenomena during the speech production that are ignored by the linear approach [6-8]. Therefore, a reasonable assumption is that a complete understanding of the dynamics of the sound signal may be the missing element in developing successful models of speech formation. There is growing interest towards developing speech synthesis and recognition techniques based on nonlinear dynamics of speech [9]. One class of methods tries to refine the classical linear methods, for instance the two-mass model, by adding nonlinear chaotic functions to the vocal cord model [10]. Another class of methods employs concepts from nonlinear dynamical theory to model the nonlinear dynamics of speech [11, 12]. For instance, a model of speech synthesis is created by embedding the one-dimensional speech time-domain waveform into an appropriate higher-dimensional space [13]. The nonlinear phenomena of the speech production, such as the degree of turbulence in speech sounds, are modeled and analyzed using fractals [14, 15]. Although the nonlinear character of speech phenomenon is widely accepted, an issue still under investigations is whether the speech time-varying signal is typically chaotic. The classical method for chaos identification is to calculate the invariants of the chaotic systems, i.e., Lyapunov exponent and correlation dimension [11, 13, 16]. Other studies try to avoid the numerical evaluation of the invariants and employ specific methods for short data sets. The results report the existence of chaos in speech [17, 18]. In a previous work, we have started to investigate the possibility of synthesizing speech phonemes starting from the dynamic model of the cognition level [19]. In the present paper we have continued to investigate the nature of phonemes dynamics at a deeper level of the acoustic signal. The remainder of the paper presents the following: The investigation of the global dynamics for five vocal phonemes and the extraction of elemental patterns; Possibility of modeling the phonemes elementals by chaotic harmonic modulation; Simulation example of phonemes synthesis, the conclusions, and the future work.

## 2 Nomenclature

$x(t)$  – time series,  $\Delta t$  – sampling time,  $d$  – embedding dimension,  $D$  – dimension of the compact manifold,  $LE$  – Lyapunov exponent,  $f_m$  – modulating frequency,  $f_c$  – carrier frequency.

The purpose is to investigate the presence of the chaotic fluctuations in five vowels /a/, /e/, /i/, /o/, and /u/ pronounced by the same two persons, a male and a female, respectively. The vocal sound data were sampled at 96 kHz with 16 bits, and the waveforms are depicted in Fig. 1. It's interesting to observe a similar dynamics of the same vocal phoneme as uttered by the two persons, and in the same time a specific vocal "fingerprint" for each of them.

The chaotic behavior of speech phonemes was analyzed using phase space reconstruction techniques [20, 21]. The purpose was to generate several different scalar signals from the original sound signal  $x(t)$ . The embedding theorem was used, considering the embedding vector

$$X(t_i) = [x(t_i), x(t_i + \Delta t), \dots, x(t_i + (k - 1)\Delta t)], \quad (1)$$

where  $\Delta t$  is the sample time, and  $k = 1 \dots d$ , where  $d$  denotes the embedding dimension. The reason was to have samples from the original sound signal  $x(t)$  delayed by multiples of  $\Delta t$  and obtain the reconstructed  $d$ -dimensional space. According to the embedding theorem,  $d \geq 2D + 1$ , where  $D$  is the dimension of the compact manifold containing the attractor, and  $1$  is an integer, selected 1 in our experiments. Following this method, from sampled time series of speech phonemes the dynamics of the unknown speech generating system could be uncovered, provided that the embedding dimension  $d$  was large enough. In practice, the difficult problem is choosing the optimal length of the time series and the optimal time delay.

For a convenient exploration of the reconstructed phase-space we constructed the following three-dimensional map:

$$\begin{aligned} x &= x(t) \\ y &= x(t - k) \\ z &= x(t - 2k) \end{aligned} \quad (2)$$

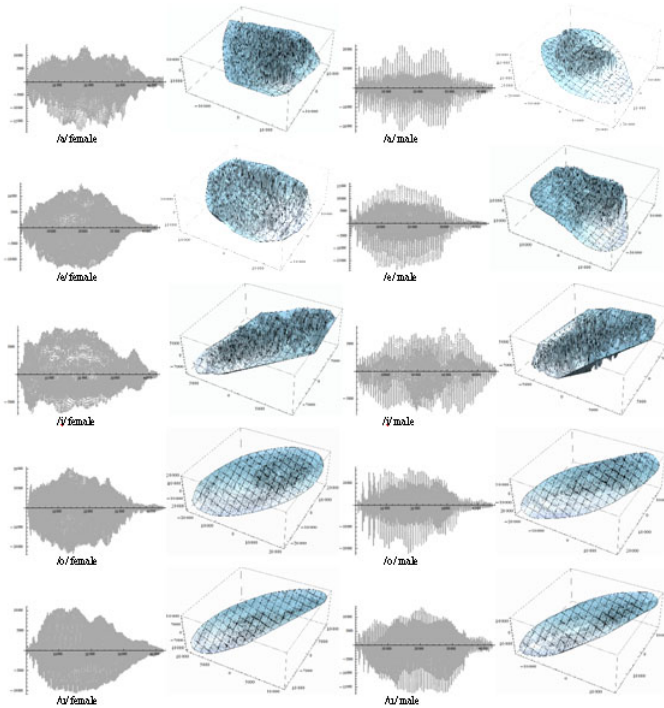
The three-dimensional attractors, computed according to (2) where  $k = d$ , are depicted in Fig. 1 along the waveforms of the phonemes.

The classical approach to analyze the dynamics of the system is through Lyapunov spectrum. Computing Lyapunov exponents (LEs) of the sampled time series is a difficult task and the results are much less precise than in the case the system's dynamics is known. The literature provides several approaches [11,17,22–26]. By definition, LE is a quantitative measure of the sensitive dependence on the initial conditions, and is an indication of the rate of convergence or divergence of initially close trajectories in the phase space. If LE is positive then the system is chaotic and the trajectories get farther apart, usually the difference increasing exponentially. If LE is zero then the attractors are regular. For LE negative the orbit attracts to stable fixed point. The dimension of the state space determines the number of different LEs that form the Lyapunov spectrum. In the case of experimental time series, it suffices to determine the largest LE in order to have an image of the system's dynamics. The trajectory tracing method was used according to [22, 26]. In real cases of limited amount of noisy data, the algorithm can provide only an estimate of the largest LE, but this is considered to be a sufficient indication of the system's dynamic behavior. However, the calculation parameters have to be carefully determined. The mutual information and the false nearest neighbor methods were used to determine the embedding delay and the optimal embedding dimension [27, 28]. Similar results were obtained using the Visual Recurrence Analysis software package [29] and the programs implemented in [26]. In order to illustrate the process, the average mutual information in dependence on the embedding delay is shown in Fig. 2, for the sampled data of phoneme /a/, for both female and male cases. The optimal time lag can be estimated for the point where the mutual information reaches its first minimum. For the same phonemes data, the percentage of false nearest neighbors (FNN) in dependence on the value of  $d$  can be seen in Fig. 3.

The principal additional input parameters of the algorithm are the maximum length scale and the minimum length scale, corresponding to the estimate of the length on which the local structure of the attractor is no longer iterated, and the maximally allowed angle separation between each successive length element. The results obtained for the phonemes under study are presented in Tab. 1.

**Table 1.** LEs estimation of vocal phonemes

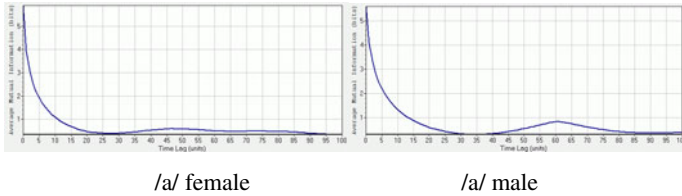
Phonemes/ Vocals	Sample length	Optimal embedding dimension	Optimal time delay	LE $\approx$
/a/, f	43312	17	27	34
/a/, m	43304	19	34	42
/e/, f	43056	19	49	256
/e/, m	42984	14	59	16
/i/, f	42624	18	50	207
/i/, m	42240	9	19	167
/o/, f	43680	9	39	27
/o/, m	45120	15	48	53
/u/, f	42432	19	50	30
/u/, m	42624	10	62	54



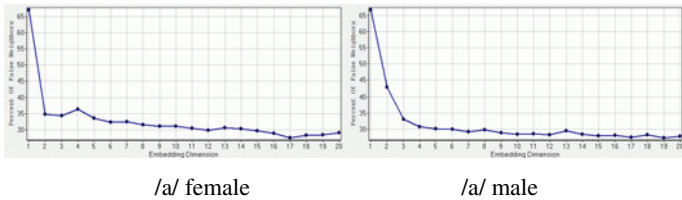
**Fig. 1.** The waveforms of the vowels /a/, /e/, /i/, /o/, /u/ and their dynamic behavior according to (2)

The maximum values obtained for LEs are rather small positives and proved the chaotic characteristic of natural human speech. This was an expected result, taking also into account other previous reports [10, 11, 16, 17]. Considering that the used

algorithm was highly dependent on the estimation of the input parameters the results might have been affected by errors. However, we concluded that speech phenomenon may be classified in the domain of chaotically deterministic systems with small positive values of the LE. This may be an indication of a slow divergence rate from the equilibrium state, suggesting a possible controlled chaos phenomenon.



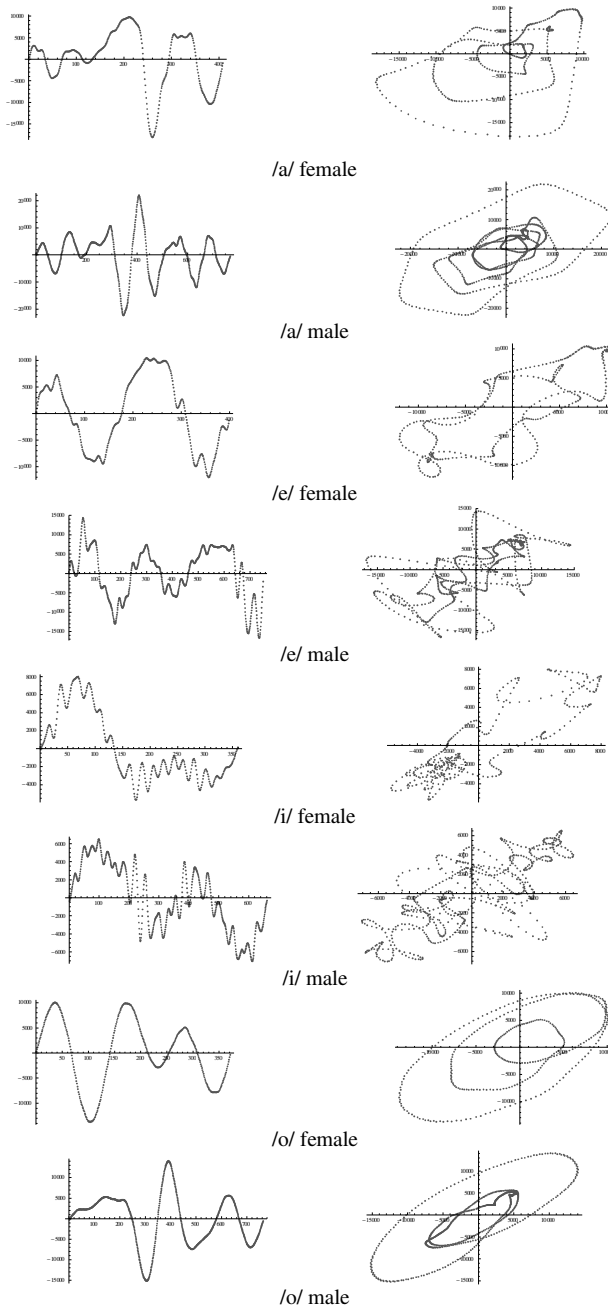
**Fig. 2.** The optimal time delays result in the point where the average mutual information reaches the first minimum (27 and 34 respectively)



**Fig. 3.** The percentage of FNN should drop to zero for the optimal global embedding dimension  $d$  (17 and 19 respectively)

However, this result was not considered yet very conclusive when taking into account the elements for devising a model of phoneme synthesis. Have all the elements that concur in the ensemble model a chaotic nature or only some of them? In the present work, the purpose was to investigate this problem in the case of chaotic formant synthesis, and try to identify the exact elements that may be the source of chaos. We analyzed the phonemes dynamics at a deeper level. The phonemes waveforms such as those recorded in Fig. 1 have a clear periodic nature. For each phoneme, the entire signal is composed of a series of repetitive patterns that show slight variations of the signal parameters. However, among all the variations, a typical pattern or elemental can be identified in each phoneme as presented in Fig. 4. The elemental waveforms are depicted along with the corresponding embedding space. The dynamics properties of these elemental patterns were also investigated, although with a higher expectancy of errors due to the smaller length of the sampled data. The same method as described above was used. The global results indicated a very small positive value of the  $LEs$ , as presented in Tab. 2.

It was interesting to note that if the elemental sample for each phoneme was concatenated to itself in order to reconstruct the phoneme to the original time length, the resulted sound was not as natural as the original. This was an indication that the elementals might be affected by slight chaotic changes one after another as they are chained in time to form the phoneme.



**Fig. 4.** The vocal elemental patterns and the corresponding embedding space

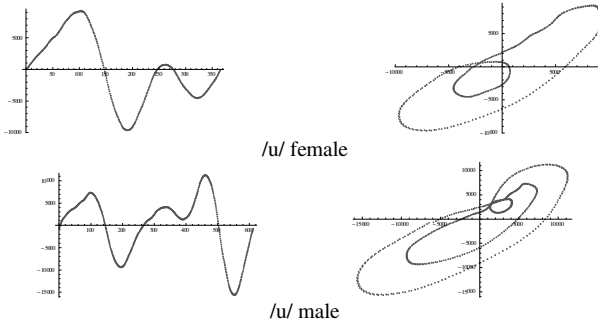


Fig. 4. (Continued)

Therefore, the suggested approach was to model the elementals by a classic harmonic modulation and then influence the parameters by chaotic signals during the process of concatenating the elementals in time for phoneme synthesis. The expression used for modeling the elementals is:

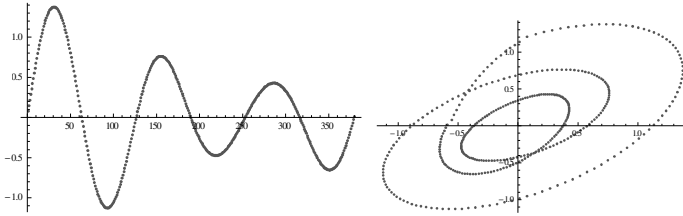
$$F(t) = (m_i \text{Cos}(2 \pi f_m t) + m_a) \text{Sin}(2 \pi f_c t), \tag{3}$$

where  $m_i$  and  $m_a$  are amplitude parameters,  $f_m$  is the modulating frequency and  $f_c$  is the carrier frequency.

Table 2. LEs estimation of vocal elementals

Phonemes/ Vocals	Sample length	Optimal embedding dimension	Optimal time delay	LE $\approx$
/a/, f	408	18	13	0.5
/a/, m	769	18	18	0.5
/e/, f	399	20	13	0.6
/e/, m	751	19	19	0.2
/i/, f	360	9	12	0.1
/i/, m	662	11	11	0.1
/o/, f	371	8	11	0.2
/o/, m	764	15	25	0.1
/u/, f	365	18	13	0.3
/u/, m	612	7	14	0.8

In Fig. 5, a modeling example is given for the elemental of phoneme /o/, with the values  $m_i = 0.5$ ,  $m_a = 0.91$ ,  $f_m = 0.019$ , and  $f_c = 0.079$ . The signal “/o/ female” from Fig. 4 was taken as reference for comparison. In order to obtain a similar time series as in the real case an iteration map was computed. The modeled elemental contains 380 iterates in comparison with 371 of the original. Examining both figures, a good similarity of the signals shape could be observed. Also, the similarity in dynamics could be confirmed by comparing the embedding space from both figures.



**Fig. 5.** The modeled elemental of phoneme /o/ according to (3) and the corresponding embedding space

It was interesting to note that the modeled elemental of Fig. 5 was concatenated in time, in the same way as described above, and quite good similitude was observed. However, the resulting sound of the phoneme was also not very natural, resembling the previous experiment when the real elemental patterns was used. This suggested that the impression of naturalness is not given by the phoneme’s elemental alone, but by the perception in time of the slight variations of the elementals in succession. Hence, the working hypothesis was that this slight variation is of chaotic nature.

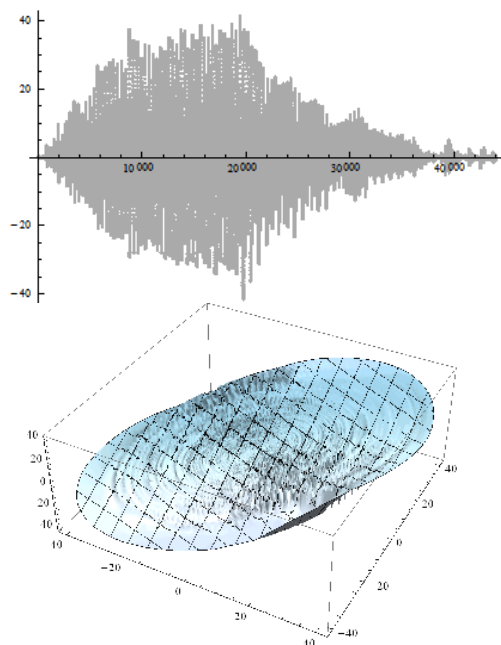
In the next phase, the values of  $m_i$ ,  $m_a$ , and  $f_m$  were taken as parameters and were controlled by a chaotic source. Different chaotic maps were used, and one acceptable result was in the case of the sine map:

$$X_{n+1} = A \sin(2\pi X_n), \tag{4}$$

where  $A = 1$  for the first 25 iterates, and  $A = 1.051$  for the rest of 91 iterates. The entire synthesis process took  $116 \times 380 = 44080$  iterates and is exemplified for phoneme /o/ in Fig. 6. The three-dimensional attractor, according to (2), is also presented. These results can be compared with those of Fig. 1 for “/o/ female.” A quite good similitude of the dynamics can be observed. Also, the signal sound was quite acceptable, creating a better impression than the simple concatenation of the elementals as described before. The maximal LE estimated value was 15, for optimal embedding dimension 3 and time delay 32.

In conclusion, the purpose was to study the possibility of phoneme synthesis starting from a dynamic model based on chaotic modulation of the phoneme’s elementals. The natural phonemes dynamics was analyzed and chaos was found in the captured time-series. The dynamic behavior, as studied for the phonemes time-series indicated a typical behavior for controlled chaos phenomena. Next, the phonemes were analyzed in terms of elementals that repeat themselves with slight variations along the signal length. The presence of chaos in the time-series suggested the possibility of modeling the elementals by a classical harmonic modulation process having the parameters influenced by chaotic signals. The model proved to be simple and effective, and the experimental results were rather satisfactory, encouraging further research in this direction. Different other models for elementals synthesis, such as Chebyshev polynomials and wavelets may be investigated in conjunction with other chaotic sources. This will be a theme for a future work.





**Fig. 6.** The modeled phoneme /o/ and the corresponding three-dimensional attractor

**Acknowledgments.** This work was supported by the Research Center of the “Politehnica” University of Timisoara.

## References

1. Jurafsky, D., Martin, J.H.: *Speech and Language Processing: An Introduction to Natural Language Processing*. In: *Computational Linguistics, and Speech Recognition*. Prentice-Hall (2000)
2. Holmes, J., Holmes, W.: *Speech Synthesis and Recognition*, 2nd edn. Taylor & Francis, N.Y (2001)
3. Mitkov, R. (ed.): *The Oxford Handbook of Computational Linguistics*. Oxford Univ. Press (2005)
4. Styger, T., Keller, E.: Formant synthesis. In: Keller, E. (ed.) *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*, pp. 109–128. John Wiley, Chichester (1994)
5. Ishizaka, K., Flanagan, J.L.: Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Systems Technical Journal* 51, 1233–1268 (1972)
6. Teager, H.M., Teager, S.M.: Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract. In: Hardcastle, W.J., Marchal, A. (eds.) *Speech Production and Speech Modelling*. NATO Advanced Study Institute Series D, Bonas, France, vol. 55 (July 1989)
7. Maragos, P., Potamianos, A.: Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition. *Journal of Acoustical Society of America* 105(3), 1925–1932 (1999)

8. Nomura, H., Funada, T.: Sound generation by unsteady flow ejecting from the vibrating glottis based on a distributed parameter model of the vocal cords. *Acoustical Science and Technology* 28(6), 392–402 (2007)
9. Faundez-Zanuy, M., et al.: Nonlinear Speech Processing: Overview and Applications. *Control and Intelligent Systems* 30(1), 1–10 (2002)
10. Koga, H., Nakagawa, M.: A Chaotic Synthesis Model of Vowels. *Journal of the Physical Society of Japan* 72(3), 751–761 (2003)
11. Pitsikalis, V., Kokkinos, I., Maragos, P.: Nonlinear analysis of speech signals: Generalized dimensions and Lyapunov exponents. In: *Proc. European Conf. on Speech Communication and Technology-Eurospeech 2003*, pp. 817–820 (September 2003)
12. McLaughlin, S., Maragos, P.: Nonlinear methods for speech analysis and synthesis. In: Marshall, S., Sicuranza, G. (eds.) *Advances in Nonlinear Signal and Image Processing*, vol. 6, p. 103. Hindawi Publishing Corporation (2007)
13. Banbrook, M., McLaughlin, S., Mann, I.: Speech characterization and synthesis by nonlinear methods. *IEEE Trans. Speech Audio Process.* 7(1), 1–17 (1999)
14. Pitsikalis, V., Maragos, P.: Filtered dynamics and fractal dimensions for noisy speech recognition. *IEEE Signal Processing Letters* 13(11), 711–714 (2006)
15. Pitsikalis, V., Maragos, P.: Analysis and classification of speech signals by generalized fractal dimension features. *Speech Communication* 51(12), 1206–1223 (2009)
16. Koga, H., Nakagawa, M.: Chaotic and Fractal Properties of Vocal Sounds. *Journal of the Korean Physical Society* 40(6), 1027–1031 (2002)
17. Tao, C., Mu, J., Xu, X., Du, G.: Chaotic characteristic of speech signal and its LPC residual. *Acoust. Sci. & Tech.* 25(1), 50–53 (2004)
18. Liu, X., Povinelli, R.J., Johnson, M.T.: Detecting Determinism in Speech Phonemes. In: *Proc. of IEEE Signal Processing Society 10th Digital Signal Processing Workshop*, pp. 41–46 (2002)
19. Crisan, M.: Upon phoneme synthesis based on dynamic semantic modeling. In: *Proc. IEEE 2nd International Conference on Software Engineering and Data Mining (SEDM)*, pp. 467–472 (2010)
20. Ruelle, D.: *Chaotic Evolution and Strange Attractors*. Cambridge University Press (1989)
21. Small, M.: *Applied Nonlinear Time Series Analysis. Applications in Physics, Physiology and Finance*. World Scientific Publishing Co.(2005)
22. Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A.: Determining Lyapunov exponents from a time series. *Physica* 16D, 285–317 (1985)
23. Eckman, J.-P., Oliffson Kamphorst, S., Ruelle, D., Ciliberto, S.: Liapunov exponents from time series. *Phys. Rev. A* 34(6), 4971–4979 (1986)
24. Brown, R., Bryant, P., Abarbanel, H.D.I.: Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Phys. Rev. A* 43(6), 2787–2806 (1991)
25. Kinsner, W.: Characterizing chaos through Lyapunov metrics. In: *Second IEEE International Conference on Cognitive Informatics (ICCI 2003)*, p. 189 (2003)
26. Kodba, S., Perc, M., Marhl, M.: Detecting chaos from a time series. *Eur. J. Phys.* 26(1), 205–215 (2005)
27. Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33(2), 1134–1140 (1986)
28. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase space reconstruction using a geometrical construction. *Phys. Rev. A* 45(6), 3403–3411 (1992)
29. Kononov, E.: *Visual Recurrence Analysis Software Package, Version 4.9* (accessed 2011), <http://nonlinear.110mb.com/vra/>
30. Vogels, T.P., Rajan, K., Abbott, L.F.: Neural networks dynamics. *Annual Review of Neuroscience* 28, 357–376 (2005)
31. Rabinovich, M.I., Varona, P., Selverston, A.I., Abarbanel, H.D.I.: Dynamical principles in neuroscience. *Reviews of Modern Physics* 78(4), 1213–1265 (2006)

# Image Processing Scheme for Mobile Communication Equipment

Ye Liang

Department of Computer Science, Beijing Foreign Studies University, Beijing, China  
liang\_ye@schu.com

**Abstract.** In order to execute more sophisticated multimedia computing tasks, leverage potential computing resources, and execute tasks on external devices, a Mobile Peer-to-Peer (MP2P) framework is proposed in this paper. The MP2P framework technique is composed of group management technique, resource management technique, and pipe technique, and it could help mobile communication units sharing resource with greater efficiency and speed. The experimental results show that the services running on MP2P framework have better efficiency than on traditional client / server model framework. Since the nodes connected by left branch are better than right branch, the left branch provides a good means to get an optimal route for mobile communication units.

**Keywords:** image processing scheme, Mobile P2P infrastructure, mobile communication equipment.

## 1 Introduction

In mobile computing field, the handheld mobile devices, such as cellular phone, Personal Digital Assistant (PDA), notebook PC with wireless communication module, could help us to get information we want. At the same time, a large number of companies begin to build resource servers and provide all kinds of resource to their customers, especially the resource which may be used outdoors. When the tourists make a journey, for example, they usually come to be perplexed in a strange land, although they have been ready for this trip. In this case, they could use the cellular phone taken with them to download the information correlative with their puzzle from the GIS server [1, 2, 3].

However, the communication based on client / server model could bring on high load to central server, and the costs of air transmission are so high that there are few people could accept it. But we take note of that the message, required by users coming from the same group with wireless communication equipments, usually focuses on a certain area. So, we could use the mobile peer-to-peer (MP2P) technology to resolve this confliction. MP2P network can let messages be transferred in the group instead of in the whole communication system, and this could save the messages in the group

and meet the need of other users through storage and transmitting. So the users could reduce the rate of communication and cut down the fee for using mobile communication devices.

This paper is organized as follows. In Section 2, we briefly summarize related work. In Section 3 we explain the conception of MP2P. In Section 4, we present the technique of organization a MP2P network. Experimental results about MP2P network are reported in Section 5. At last, we give some conclusions and future work.

## 2 Related Work

Numerous efforts are being made in the direction of using handheld mobile devices for the purposes of sharing information. Some technologies simply store the shared information in the database server, and distribute these resources through browser / server (B/S) mode or client / server (C/S) mode to users' handheld devices. Other technologies plan to realize the function of sharing resources by means of the third generation mobile phone. By examining a few related applications and concepts, we shall see how MP2P makes the sharing technology one step further.

Now, most people get information using the traditional network technique, which base on B/S or C/S network model. In this model, the users must have mobile devices which can communicate with remote server directly or get connection with the local network, such as Bluetooth network or Ad Hoc network, around them, so that the mobile device could interaction with remote server via the local network. But there are two disadvantages to this technique: On the one hand, the bandwidth of wireless link is a major performance bottleneck. The more users want to get the similar resource from server at the same time, the more response time is needed for each user. On the other hand, wireless communication is very expensive, and the burden of fee would fell to the users after they use this kind of application for a long time. [4, 5, 6]

There are some technologies designed for the third general mobile phone. These technologies mainly concentrate on how to use P2P technique which has been used widely at internet in the mobile communication field. But yet, P2P technique can't be used on the mobile phone of general packet radio service (GPRS) networks or CDMA 1X networks, often called 2.5G networks generation, in which an IP address, unlike the third generation mobile phone, can't be distributed for each, so the research on P2P for mobile phone only has theoretic value [7].

## 3 The Conception of MP2P

In the interest of the users, we propose a new technique, MP2P technique, based on mobile unit in this paper. Using the MP2P technique, we can get the resource we need with shorter response time and lower fee via our mobile phone which we have had.

MP2P network is a virtual network built on basic communication network, and MP2P is a technique with which handheld mobile devices could communicate with each other by peer to peer method. As we all know, the mobile phone we use could not get an IP address for itself, but it has a fixed phone number with it, so we can recognize it accurately. At the same time, as the mobile phones are updated incessantly, most have Bluetooth module or BREW module in them. That gives me a method to make the handheld devices to connect with each other directly, or through a pipe. If the handheld mobile devices we are using have BREW platform of QUALCOMM or J2ME platform of SUN, we can design platform-independent and language-independent program running on them to let them share the resource in MP2P network. Because these resources are only exchanged at inner of local MP2P network which is created by local creating network modules fixed in the devices of one group, and need not be supported by 2.5G network, the communication time and fee is acceptable.

With technique developing, there are more and more techniques could be used on handheld mobile devices as running technique for MP2P network in mobile computing field. They are platform-independent, language-independent, so they could be applied widely, and the MP2P technique could have a broad prospect with them.

## 4 Organizing a MP2P Network

In a mobile computing environment, handheld mobile devices is always moving; signal might break off frequently; the bandwidth of radio links is narrow; and finally the bandwidth of air transmission is asymmetrical. As a result, we need to design a new network organization technique to suit for wireless environment. It consists of group management technique, resource management technique, and pipe technique.

### 4.1 Group Management Technique

Group management technique includes creating group, entering group, and exiting group.

- Creating group

At first, all the handheld mobile devices use Bluetooth module or Ad hoc module to connect with each other and buildup a local MP2P network, and all of them makeup a basic group. When a task, for instance, downloading a image all about them, is brought forward in the basic group, a device which could do it immediately is selected to buildup a new sub-group by the mobile agent running in the MP2P network, and this device is called leader of the new group. The process of creating group could be figured as Fig.1

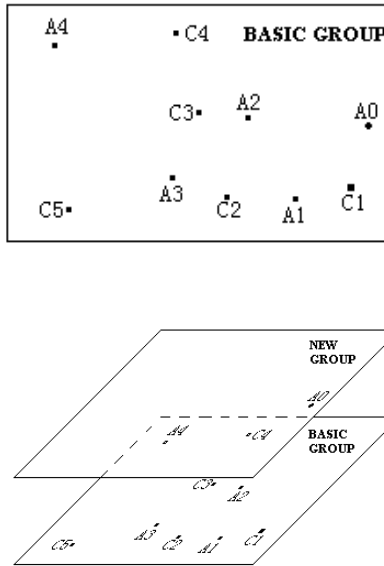


Fig. 1. Device A0 is selected to build up a new group

- Entering group

After the leader device builds up a new group, the devices which maybe use the same multimedia resource join into the group one by one. But how to decide whether a device needs to enter the group? In order to explain the regulation in detail, we put forward some definitions.

**Definition 1.** Relativity: Suppose there are two devices named  $N_A$  and  $N_B$  respectively, the aggregate of attributes for  $N_A$  is  $A_{\text{attribute}}$ , and for  $N_B$  is  $B_{\text{attribute}}$ , the correlation coefficient could be expressed by  $R_{AB}$ , then

$$R_{AB} = \frac{A_{\text{attribute}} \cap B_{\text{attribute}}}{A_{\text{attribute}} \cup B_{\text{attribute}}}$$

According to the definition,  $R_{AB} \in [0, 1]$ , if the value of  $R_{AB}$  is high, the attributes of  $N_A$  and  $N_B$  are similar.

**Definition 2.** Threshold: When a device wants to join into a group, it need calculate the correlative degree with the devices of the group. If all the values are higher than a certain value, it will be allowed to enter the group, and this certain value is called Threshold (T).

So, if the leader device  $N_A$  creates a new group G, and all the devices in the group G are denoted as  $N_G$ , that is  $|N_G| \geq 1$ . When a new device  $N_B$  want to enter the

group, if  $|N_G| > 1$ , and  $|\{N_c | R_{BC} \geq T, N_c \in N_G\}| \geq 2$ , or  $|N_G| = 1$ , and  $R_{AB} \geq T$ , the new device could enter this group.

- Exiting group

If a device doesn't need the sharing multimedia resource again, and it wants to get a new kind of resource from another group, it could apply to exit current group. If the leader device needs to exit the group, the mobile agent of managing group must select a new leader for the group.

## 4.2 Resource Management Technique

After creating a group, the mobile agent needs to control the distribution of resource. To describe the resource management technique better, we put forward some definitions and theorems.

**Definition 3.** Atomic resource message: It is also called atomic message (AM) for short. It is the minimum block of message which could be cut into, but still could be used in the future. Only AM could be allowed to transfer in the group at MP2P network. We use the mark  $R^*$  to denote it.

**Definition 4.** Combinative resource message: It is also called combinative message (CM) for short. It makes up of a serial of AM. The method of combination concludes combination on different dimension, combination on the same dimension, and both of above. We use the mark  $R$  to denote it. Then  $R = n R^*$  ( $n > 1$ ).

**Definition 5.** Resource binary tree (RBT): It is used to record a certain resource of group. In binary tree, the node represents the handheld mobile device, and the branch represents the two devices with a direct communication link. The leaf of tree represents the AM of group at a certain device.

In RBT, there is a branch between two nodes. This represents they could share resource with greater efficiency and speed, and the nodes connected by left branch are better than the nodes connected by right branch. According to the definition, only the AM could be represented by RBT, but we usually need CM as we work. So we must ensure that the AM and CM could transform into each other.

**Theorem 1.** In MP2P network, CM could be combined by AM.

Proof. Let A, B, C, D are AMs, A, B at the same dimension, C, D at the different dimension, and A, B could not be combined with C, D,

$$\text{Then } R_{AB} = R_A^* + R_B^*, \quad R_{CD} = R_C^* \cup R_D^*$$

$$\text{So, } R_{ABCD} = R_{AB} + R_{CD} = (R_A^* + R_B^*) + (R_C^* \cup R_D^*)$$

That is the AMs could transform into a CM.

Proof is over.

**Theorem 2.** In MP2P network, AM could be transformed by CM and the other AMs.

Proof. Let A, B, C, D are AMs,  $R_{ABCD}$  is a CM, which is combined by A, B, C, D. With theorem 1, we know:

$$R_{ABCD} = (R_A^* + R_B^*) + (R_C^* \cup R_D^*)$$

If we want to get A or B,

$$\text{Then, } R_A^* = R_{ABCD} - R_B^* - (R_C^* \cup R_D^*)$$

$$\text{or } R_B^* = R_{ABCD} - R_A^* - (R_C^* \cup R_D^*).$$

We obtain that A or B could be transformed.

If we want to get C or D,

$$\text{Then, } (R_C^* \cup R_D^*) = R_{ABCD} - R_A^* - R_B^*$$

$$\text{That is } R_C^* = (R_{ABCD} - R_A^* - R_B^*) / R_D^*$$

$$\text{or } R_D^* = (R_{ABCD} - R_A^* - R_B^*) / R_C^*$$

We obtain that C or D could be transformed.

So, AM could be transformed by a CM and the other AMs.

Proof is over.

Using the RBT, according to theorem 1 and theorem 2, we can also use binary tree to buildup a route from the handheld mobile device which requests the resource to the handheld mobile devices which have the multimedia resource or part of the multimedia resource meeting the request. For example, a device requests the CM: AUB+ CUE, and it uses the RBTs of AM: A, B, C, D and binary tree method to create a route of getting these AMs. Fig.2 shows how to create the binary tree.

### 4.3 Pipe Technique

**Definition 6.** Pipe technique: In MP2P network, pipe technique is that a device could get the multimedia resource from another group directly.

In MP2P network, to get multimedia resource also needs the pipe technique. Pipe technique is a core technique of computing field, and we will introduce it in next paper in detail.

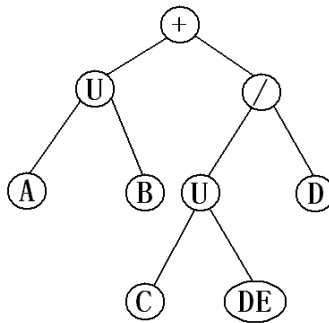
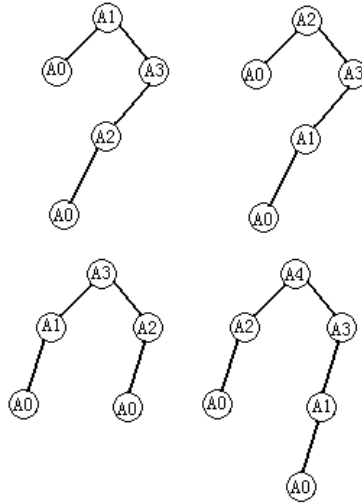


Fig. 2. Create a binary tree to get resource AUB+ CUE

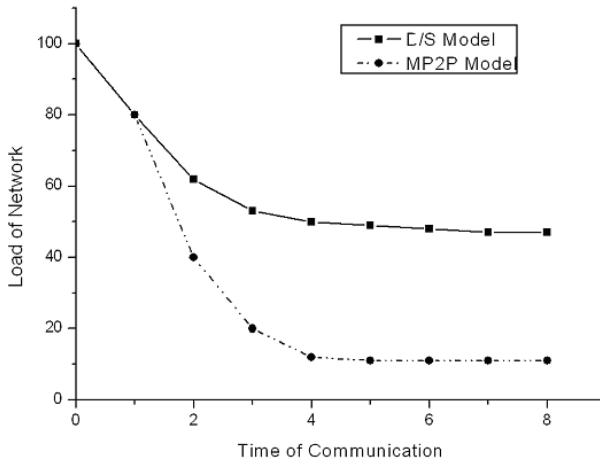


## 5 Experimental Results



**Fig. 3.** Mobile devices index the image resource of server

In journey, the image resource is a sort of information the tourists need frequently. Fig 3 respectively shows the handheld mobile devices of tourists index the same image resource using RBT.



**Fig. 4.** Compare the fee between B/S and MP2P model

At first, the image resource only is saved on the server, so any image must be got from it. In MP2P network, we suppose the server is a number of the group, and call it

A0, A1, A2, A3, A4 all could get the image from A0 through the index of RBT when they need. So, no matter which method is used to get the image, the fee is equal. However, after anyone of them gets the image from A0, the others don't need get the image from A0, and they could get it via the MP2P network. In Fig.4, the B/S model and MP2P model are compared at the data quantity of communication between MP2P network and GPRS networks or CDMA 1X networks, and this figure also stands for the fee for communication.

## 6 Conclusions

With MP2P, any group working in the field that requires a collaborative channel for achieving a goal could be brought to success with less time and costs. So, the MP2P technique has a broad prospect in the image processing field.

**Acknowledgment.** This paper is supported by the Fundamental Research Funds for the Central Universities (No.2010XJ017). Without this help, this work would never have been completed.

## References

1. Liang, Y.: GG TPR-tree Indexing Method for Grouping Moving Objects. In: The 2011 International Conference on Computers, Communications, Control and Automation, pp. 375–378 (2011)
2. Liang, Y.: An Efficient Indexing Maintenance Method for Grouping Moving Objects with Grid. In: The 2nd International Conference on Innovative Computing and Communication and 2011 Asia-Pacific Conference on Information Technology and Ocean Engineering, pp. 313–316 (2011)
3. Liang, Y.: Research on Cache Management for Mobile Learning Devices. In: Zhou, M., Tan, H. (eds.) CSE 2011, PART II. CCIS, vol. 202, pp. 155–162. Springer, Heidelberg (2011)
4. Liang, Y.: Resource Interactive Routing Optimization Based on Genetic Algorithm. In: International Conference on Wireless Communications, Networking and Mobile Computing, pp. 874–877 (2010)
5. Liang, Y.: MP2P Based on Social Model to serve for LBS. In: International Conference on E-Business and E-Government, pp. 457–460 (2010)
6. Liang, Y.: Mobile Intelligence Sharing Based on Agents in Mobile Peer-to-Peer Environment. In: The Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 667–670 (2010)
7. Liang, Y.: Mobile Peer-to-Peer Technology Used in Water Recourse Management. In: The International Conference on Challenges in Environmental Science and Computer Engineering, pp. 93–96 (2010)

# Realization of a Multimedia Conferencing System Based on Campus Network

Wang Yun

School of Educational Technology and Communication, Shanxi Normal University, Linfen,  
Shanxi, China  
wyspn@126.com

**Abstract.** The multimedia conferencing system is a kind of multimedia communication system, which combines computer technology, communication & network technology, and micro-electronic technology. A multimedia conferencing system based on campus network is designed and realized in this essay. It can fully make use of the convenient educational resources in the campus network, overcome traditional difficulties through TV transmitting information and fulfill real time interactive meeting, teaching, multipoint communication, and at the same time records the meeting & teaching materials as courseware for replaying.

**Keywords:** Multimedia conferencing system, Multimedia communication, Campus network.

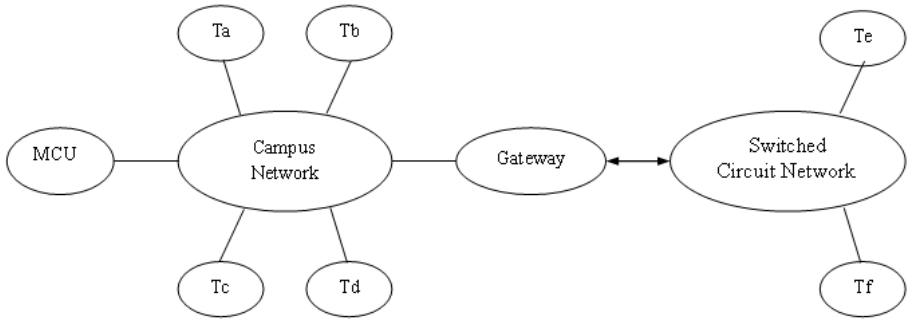
## 1 Introduction

A common problem of many universities is that thousands of students want to take some certain conferences or courses but only a few can actually take them owing to the shortage of teachers. The traditional resolution was videoing and then broadcasting through TV. This used to play an important part in television education, but it can't support the interaction between the teacher and the students and the discussion among the classrooms. With the network becoming more and more popular, more and more people are receiving the concept of network education rather than TV. The multimedia conferencing system is a kind of multimedia communication system, which combines computer technology, communication & network technology, and micro-electronic technology. The multimedia conferencing system based on campus network, which is designed and realized in this essay, can fully make use of the convenient educational resources in the campus network, overcome traditional difficulties through TV transmitting information, solve the problem of shortage of teachers of some popular courses and fulfill real time interactive meeting, teaching, multipoint communication, and at the same time records the meeting & teaching materials as courseware for replaying.

## 2 Constituents and Standard of the Multimedia System

### 2.1 Constituents

Fig.1 is the structure of multimedia conferencing system based on campus network. It is composed of many parts as the following: the campus network, a MCU (Multipoint Controller Units), a gateway, a switched circuit network and many terminals, in fig.1, Ta, Tb, Tc and Td are H.323 terminals, Te is a H.320 terminal, Tf is a H.324 terminal.



**Fig. 1.** The structure of multimedia conferencing system based on campus network

### 2.2 Standard

The standard of this system is H.323. It is ITU standard for real-time multimedia communications [1], and it is an umbrella recommendation that sets standards for multimedia communication across IP-based networks, such as networks based on Ethernet, fast Ethernet and token ring technologies. The recommendation deals with networks that do not provide a guaranteed QoS [1]. H.323 can be deployed wherever Internet Protocol (IP) is supported, regardless of physical network topology or protocol that includes Ethernet, ATM, FDDI, ISDN, etc. Its standard is both hardware and operating system independent, which means products can be manufactured by many different vendors for use in a variety of environments.

## 3 Gateway and MCU of the Multimedia System

### 3.1 Gateway

The gateway of multimedia conferencing system based on campus network provides connectivity between H.323 terminals and other types of terminals. A gateway allows an ISDN (H.320), POTS (H.324), or analog telephone system to call into an H.323 conference, and vice versa. The H.323 standard requires that gateways provide call signaling support, control channel messages, multiplexing, audio compression, and audio transcoding. The standard of this system does not define the number of ports a gateway must support or the type of physical interface a gateway should have.

### 3.2 MCU (Multipoint Controller Units)

MCU (Multipoint Controller Units) is the linchpin of multimedia conferencing system based on campus network, it is a server of high performance and is able to realize information switching & controlling between all terminals. MCU consist of a required multipoint controller (MC) and an optional multipoint processor or processors (MP). [2] The MC component of the MCU is the conference controller. The MC handles negotiation between all terminals to determine common capabilities and controls conference resources such as multicasting. The MC does not deal directly with any of the media streams. The MP component of the MCU processes, mixes, and synchronizes the media streams that are included in the conference. The MCU function can be provided in a dedicated hardware unit on the LAN or it may be embedded within other H.323 entities. MCU of this system deals mainly with three kind of information as following.

**Video signals:** MCU of this system can directly assign video signals, if some people wants to speak in any a terminal, their pictures can be carried MCU and can be switched by MCU to any other terminals which are linked MCU.

**Audio signals:** The audio processor of MCU fulfills audio signals of this system, which is composed by code converter and audio mixer. When one meeting spot has speaking signal, MCU of the system will switch Audio signals to other spots, while several spot have speaking signal at the same time, MCU will select one of Audio signals and switch it with other meeting spots according to meeting control pattern.

**Data information :** Data information processing of the system is performed by data processor; it mainly supports electrical whiteboard, static image switching, file commuting and storage.

### 3.3 Control Mode

The system employ centralized multi-spot control manner, all its terminals have been set up connection with point-to-point, each terminal send its media stream (audio, video, data, control) to MCU, when these data are dispose by MCU, these media stream will be distributed partly or whole to every terminals.

## 4 Terminals of the Multimedia Conferencing System

Terminals are the clients in an H.323 network. They are typically video conferencing, audio conferencing, or other multimedia systems implemented by end users to communicate in real time. The H.323 standard requires that every Terminals support certain functions and codec that have been previously defined by the ITU [3]. Fig.2 shows a terminal of multimedia conferencing system based on campus network. The network interface of the terminal is described by the H.255 protocol. ITU-T Recommendation H.225 (1998): "all signaling protocols and media stream package for packet based multimedia communication systems [4]." This defines the signaling for connection establishment called RAS (registration, admission control and status) signaling. The terminals of multimedia conferencing system based on campus network are required to support the following:

H.261, H.263 and H.263+ are the ITU standards of video compression. Video codec of terminals are required to support them. G.711, G.728 and G.729 are the ITU standards for speech compression; audio codec of terminals must support them. They can performance following operators : multi-spot electrical whiteboard (T.126), multi-spot file transmitting (T.127), multi-spot share (T.128). H.245 is a protocol for controlling media between H.323 terminals, and it is used to exchange control messages such as flow control and channel management commands. Q.931 is a signaling protocol for establishing and terminating calls. RAS channel is a data stream used to communicate with a gatekeeper, and it is may used to exchange signal about call registrations, admissions, and terminations. RTP/RTCP is ETF (Internet Engineering Task Force) application level protocol that RTP uses to control and synchronize streaming audio and video. [5] It provides feedback information to the source that can be used to adapt the flow to changing network conditions.

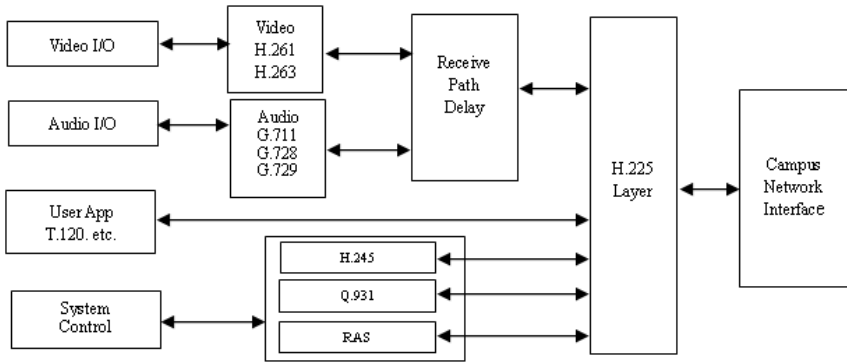


Fig. 2. The terminal of multimedia conferencing system based on campus network

Optionally, the terminal of multimedia conferencing system based on campus network may also support video, additional audio codec, T.120 data conferencing, multicast, and QoS. If a terminal supports video, the standard requires a minimum of H.261 QCIF (Quarter Common Interchange Format) support. Support for other video formats is optional.

## 5 Classrooms of the Multimedia System

There are two kinds of classrooms in the multimedia system; they are respectively entitled primary classroom and secondary classroom.

### 5.1 Primary Classroom

The primary classroom is where the teacher stays. This classroom is composed of a server, a teacher's PC, a video recorder, an electronic white board and two overhead projectors. The teacher's PC is used to play slide of the lecture, and it projects the

slide to the electronic white board. If the teacher writes something on the electronic white board, the teacher's PC will capture the written information and combine it with the slide. At the same time, the teacher's PC compresses the slide/written information and sends it to the server for recording, and sends the information to classrooms for broadcasting with multicast.

## 5.2 Secondary Classroom

The secondary classroom is the classroom without the teacher. It is made up of a server, a video recorder, an overhead projectors and an electronic white board. The server receives the information from the primary classroom, and projects the video information on the white wall. The students in this classroom can discuss with the teacher in the primary classroom each other.

## 6 Conclusion

The multimedia conferencing system based on campus network is designed and realized. It will play an important role in fulfilling real time interactive meeting, teaching and learning, multipoint communication, and at the same time records the contents as courseware for replaying.

## References

1. Gibson, J.D., Berger, T., Lookabaugh, T., Lindbergh, D., Baker, R.L.: Digital Compression for Multimedia Principles & Standards. Morgan Kaufmann Publishers, Inc. (1998)
2. Black, L.: Emerging Communications Technologies, 2nd edn. Prentice Hall (1997)
3. Zhu, Z.: A general survey of research in standardization of e-learning technology. Open Education Research (4) (2007)
4. Passerini, K.: A Developmental Model for Distance learning Using the Internet. Computer & Education 34, 1–15 (2000)
5. Kuo, F., Effelsberg, W.: Multimedia Communication: protocols and applications. Prentice-Hall (1997)

# Study on the Wireless Access Communication of Line Condition Monitoring System of Smart Transmission Grid

Jinghong Guo<sup>1</sup>, Hao Zhang<sup>1</sup>, Yaxing Liu<sup>2</sup>, Hongbin Liu<sup>2</sup>, and Ping Wang<sup>2</sup>

<sup>1</sup> State Grid Electric Power Research Institute(SGEPRI), Jiangsu, China

<sup>2</sup>North China Grid Company Limited, Beijing, China

**Abstract.** An extensive communication system is the essential infrastructure of the Smart Grid to ensure the information exchange and resource management rapidly and efficiently. Since it will affect all areas of power grid, from generation system to transmission and distribution system, the communication system will be hybrid and involve various communication technologies, such as optical transmission, wireless techniques, advanced sensor networks, etc. In this paper, we focus the discussion on the broadband wireless access techniques used for monitoring power transmission. We propose a hybrid wireless monitoring system based on WiFi and WiMAX to complete the dual direction communications for data collection and system controlling. The system design is introduced and the challenges for system implementation are discussed. Through our work, we wish to make contributions to China Smart Grid.

**Keywords:** Wireless Access, Monitoring power transmission, Smart Grid.

## 1 Introduction

With the increasing attentions to limited energy resources and the usage of green energy, Smart Grid becomes the common expectations for most of countries' state grid system. The Smart Grid will involve various technologies, including sensing, communications, networking, security, data mining, analytics, control and management, etc. Besides utilizing the state of the art technologies, some organizations also draft or release series of standards for Smart Grid to improve the interoperability among different systems and equipments provided by worldwide vendors, such as IEEE 1547 series of standards for interconnections[1], IEEE P1901 for BPL (broadband over power line) and HOMEPLUG1.0 standard [2]. The research on Smart Grid is investigated widely and some projects or trails are deployed in United States or Europe [3][2]. In China, constructing a national Smart Grid is also involved in the plan of the government [4]. To 2020, China will be fully integrated into a strong smart grid, so that the grid resource allocation capabilities, the level of security and stability, as well as network with the power and the interaction between users to be significantly enhanced [5]. An extensive communication network is the essential infrastructure of the Smart Grid [6].



Smart power transmission system is the basic element of the Smart Grid. Monitoring the power line and substations can reflect the current status of the transmission system. With those data sampled by the monitoring system, the central management system can detect and respond the system errors more efficiently. Thus it will enhance the stability of the transmission system. Meanwhile, the distribution of power generation will be more reasonable with the information of the transmission system. Currently, there are several communication solutions for power transmission system monitoring:

- Digital Transmission using FSK, MSK, etc. This is a typical narrow band transmission technique. The data rate is about 300-19200bps and can not support multimedia applications, such as video surveillance. The typical user interface is RS232 or RS485.
- Optical Transmission. This solution can provide enough bandwidth for data transmission. The precondition is that the fiber is laid along with the power line monitored. This can not be achieved by every country, especially for those countries like China which has large rural area and varied geography.
- Wireless Transmission. In traditional, the wireless transmission is treated as a backup solution for the wired transmission. With the rapid development of wireless technologies and the popularization of the wireless network, the wireless transmission becomes a strong competitor for power transmission monitoring system. The wireless monitoring system can be divided into two types. One is transmission using public wireless access network, such as GSM/GPRS/CDMA. The other is to construct the industry proprietary network using broadband access technology, such as WiMAX (Worldwide Interoperability for Microwave Access). Compared the two solutions, the later one has distinct advantages on system maintenance, security, management, and scalability. And from the long term view, the total cost will be lower. Therefore, we believe an industry proprietary wireless network will be the trend for the wireless monitoring system of power grid.

For China power grid, the current transmission monitoring system is mainly used the narrow band digital transmission technology. 230MHz and 800MHz are the frequency resource assigned by China SRRC (State Radio Regulation Committee) for the usage of digital transmission. A few of local power grids rent the public wireless access network, such as GPRS provided by China Mobile, for power transmission monitoring with high cost. For example, the rent fee is about 42 millions USD per year for the distribution system monitoring in Yinchuan, a middle size city of China. In order to meet the requirements of Smart Grid for communications, some 3G or 4G technologies, such as WiMAX and LTE (Long Term Evolution), will be considered for the transmission monitoring system. IBM has investigated in Smart Grid research for several years. SGEPRI is the pioneer in China to explore the Smart Grid related topics. We wish to make contributions for China Smart Grid.

In this paper, we propose a hybrid wireless system for power transmission monitoring. The system uses WiFi (802.11) technology for sensing data transmission in short distance and we adopt WiMAX technology (802.16) for long distance transmission. The design constraints and potential solutions are addressed in this paper. And the challenges for system implementation are also discussed in details.

The rest of paper is organized as follows. Section II introduces the system deployment requirements for the power transmission monitoring system. The system architecture is proposed in Section III and the system design is described in details. In Section IV, we discuss the challenges of system implementation. Finally, we conclude the paper in Section V and propose the future work.

## 2 System Deployment Requirements

Considering the real situations of China power grid, we suppose the distance between two substations is averagely 200km. We set the base station at each power substation. Thus each base station should cover the radius about 100km. Considering the signal attenuation and path loss, relay station or repeater must be used in this scenario. The system deployment illustration is depicted in Fig. 1. The detail information will be given in Section III. Table 1 shows the deployment requirements for the monitoring system.

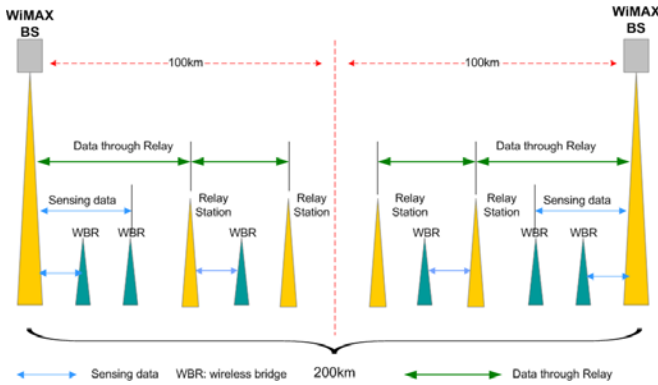


Fig. 1. System Deployment illustration

Table 1. Deployment Requirements for Monitoring System

Items	Requirements
Objects being monitored	Temperature, galloping, vibration, gradient of pole tower, video, ice and snow, etc.
Distance between substations	about 200km
Distance between two power tower	50~100m
Power supply of device on power tower	15watts
Operation spectrum	license exempt frequency band, such as 230MHz
Temperature requirement	-40~75 Celsius degree
Aggregate throughput	384Kbps
Physical interface on sensor side	Ethernet, RS232/485
Physical interface on BS side	10/100/1000 Ethernet for data 0/100 Ethernet for management
Latency requirement	10s

### 3 Hybrid Wireless Monitoring System Design

#### 3.1 System Architecture

Based on the deployment requirements mentioned above, the system architecture of the wireless monitoring system we proposed is illustrated as Fig. 2.

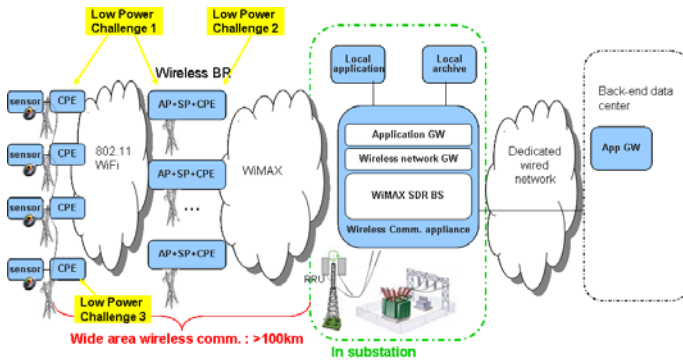


Fig. 2. System Architecture of the Wireless Monitoring System for Power Transmission

This is a hybrid wireless system for power line monitoring. Two popular wireless techniques, WiFi and WiMAX, are integrated into this system. The WiFi technique is mainly for unit sensing data transmission with short distance, while the WiMAX technique is for multiple sensors’ data transmission with relatively broad bandwidth. Software defined radio (SDR) based base station (BS) is the technical highlight of this solution which will be described in the following system design sections.

#### 3.2 System Design

1) *Sensors*: Sensors are used to acquire necessary data. They transmit data to WiFi CPE (Customer Premises Equipment) communication module through Ethernet or RS232/485 interface. Data will be encrypted at application layer. Life cycles, low power consumption and low cost are the main factors for sensors’ selection.

2) *WiFi CPE*: WiFi CPE will transmit the data through WiFi air interface to wireless bridge. Each CPE will have its IP link between WiFi CPE and AP (access point). Commercial WiFi CPE products will be considered for this module for cost saving.

3) *Wireless Bridge (WBR)*: This device consists of three parts: the AP module, the sensor proxy (SP), and the WiMAX CPE. The AP module collects the data transmitted from WiFi CPE, and the WiMAX CPE module sends out signal to the WiMAX base station (BS) located at the substation. The sensor proxy module completes the signal transforming between WiFi and WiMAX. And it also executes some extra applications, such as compression/decompression, if necessary. IP links from sensors will be optimized in WBR to save bandwidth. For this device, since it

should have multiple functions defined by the designers, the platform of the device should be flexible and easy for extension with low cost.

4) *Wireless Communication Appliance*: The basic function of this appliance is as WiMAX BS. However, it can also take the role of wireless network gateway and application gateway since we use SDR technology to implement the whole system on general IT servers with multi-cores. The SDR technique makes the system more flexible and scalable. Furthermore, the general server based platform ensures the further data processing on it. The digital signal received from RRU (remote radio unit) will be processed in SDR BS layer, the WiMAX gateway, then to the application gateway. The application gateway will do the “de-optimization” for sensor data, and decrypt the sensor data. There will be application data bus designed in application server. Through the data bus, the data can be dispatched to local database for archive, to local application server for local data processing, or to remote data center through the dedicated wired network.

At present, China SRRC has not assigned the spectrum resource for future Smart Grid yet, we just suppose the frequency bandwidth is 5MHz and the BS works at single input single output (SISO) TDD mode. Based on these assumptions, the maximal system throughput will be 6.997Mbps for downlink (DL) and 5.287Mbps for uplink (UL) if we choose 16QAM, 1/2 coding rate. Although the configuration of DL:UL ratio will affect the practical system throughput, with consideration of the unbalance characteristics of DL:UL ratio for power monitoring system, in which the uplink workload (sensing data transmission) is overwhelmed compared with the downlink controlling message, the system bandwidth we selected can meet the throughput requirements under most of cases.

But when the system is put into real deployment, considering the challenges of spectrum reuse, power constraints, and long distance coverage, more wireless system optimization work should be taken. Then it will require the modifications of PHY and MAC layer based on current 802.16e design. Our IT platform based SDR BS design can catch up with the new modifications quickly, which will be put into our future work.

## 4 Challenges for System Implementation

### 4.1 Spectrum Utilization

As the explosive growth of wireless systems deployed for machine to machine (M2M) applications, spectrum resource becomes the bottleneck for those deployments. For the Smart Grid system, it faces the same issues coming from the limited spectrum resource. Compared with the traditional fixed frequency allocations' spectrum sharing, the dynamic using of spectrum will enhance the spectrum utilization efficiency.

As we stated above, the spectrum resource for China Smart Grid has not been assigned yet. Currently, the only frequency band we can use is the 230MHz assigned for the China power grid for narrow band transmission. The total bandwidth is about 1~2M. Since the Smart Grid system is deployed steps by steps, the digital transmission system and broadband access system will be coexisted. Therefore, the

spectrum assigned shall be shared by them, shown as Fig.3. How to avoid the spectrum conflicts and improve the spectrum efficiency are big challenges for us. Although there are some prior work about spectrum sensing and optimization in cognitive radio (CR), the Grid system has specific characteristics and requirements, such as low power requirement, safety requirement, low SNR, and variant environment, etc.. The spectrum optimization algorithm shall be designed based on such requirements and constraints.

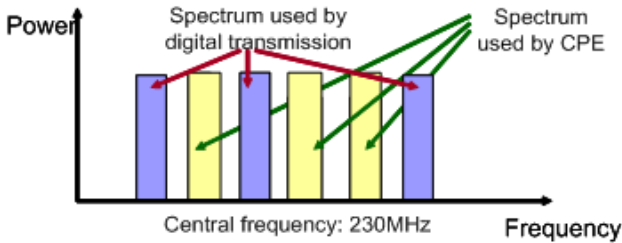


Fig. 3. Spectrum Sharing with Current Digital Transmission

## 4.2 Low Power Constraints

Our system faces several challenges in low power design in order to be a reality. A classic solar energy supply system is used to provide the energy of sensors, CPEs and wireless BRs. Up to date, the maximum power density obtainable from a modern solar cell is about  $5 - 20 \text{ mW/cm}^2$  (outdoor, sun at the zenith) that is a very limited power for our system. We explain three main challenges (as shown in Fig. 2) in the low power design of our system as follows,

- **Challenge 1:** While the solar energy supply system is effective in optimal sun conditions, the efficiency drastically falls off when a direct strong light is not granted (e.g., in presence of a partly cloudy sky, mist, morning and late hours, etc.). In our system, the sensors, CPEs and wireless BRs should work continuously on twenty cloudy days.
- **Challenge 2:** We need to find a optimal tradeoff between wireless BR's power consumption and its communication distance. If the power consumption is large, the battery is too heavy to be installed under real environments. If power consumption is small, more relays are needed.
- **Challenge 3:** With the constrained power supply, the communication distance of the CPEs should be at least 2 kilometers.

The wireless monitoring system's radio is a big energy consumer even if our solar energy system is quite efficient. Keeping the system's radio on all the time may lead to a negative energy balance. Classical WSN low power design (such as scheduling) can not be used for our system because our system is different from the classical WSN, such as hybrid communication system, routing, etc. To prolong the lifetime of the system, we need to develop a novel and efficient low power design, such as an efficient MAC protocol.

### 4.3 Relay System Design

The relays can be used to extend the coverage area of a BS and/or increase the capacity of a wireless access system. In the proposed wireless monitoring system for power line, the required wireless coverage between two substations will be 100km-300km, which is almost 10 times of the maximum coverage of one BS. Thus it is necessary to consider the relay systems for extending the coverage. On the other hand, there may be big obstacle between two substations, e.g. mountains. So, relay system can help to cover the sensors behind it.

In general, relaying systems can be classified as either decode-and-forward (DAF) or amplify-and-forward (AAF) systems. To ensure the signal quality for such a multihop system, we will consider the DAF system only. With the strict time synchronization requirement, the DAF system design for TDD mode will have more challenges than those for FDD mode. As we know, the relay system for TDD mode has just moved from the domain of an interesting research topic to standardization stage. For example, the Intelligent Relaying (IR) for UTAR TDD and 3rd generation specification is the technique in which terminals are permitted to receive and retransmit data on behalf of other users. Another important initiative is IEEE 802.16j which is developing solutions for relay-based networks that can operate with legacy 802.16e-2005. However, the relay network architecture is relatively new design and introduces many complexities within the already challenging environment of radio access networks. Many issues remain unanswered, so there is no mature product and deployment in market till now.

Compared with general wireless environment, the power line monitoring scenario in smart grid adds more challenges on coverage, power consumption and throughput, etc. Combining the traditional issues for relay system design, we summarize the design challenges as following.

- **The stability of relay system with many hops:** The power line monitoring system in smart grid will require more than 20 hops between two substations. No matter industry standard or academic research, most of the current TDD relay systems are designed for only 2-3 hops. Unlike the general ad hoc network where there will be a few redundant routes for each relay node, this wireless power line monitoring system will use chain topology. So, the stability will become a big challenge.

To make the network robust, some extra designs about automatic diagnosis will be helpful. On the other hand, the latency will be increased due to the relay. RS close to BS will have less increase than those far away from the BS. So, the timeout settings on each RS should be tuned accordingly.

- **System efficiency:** To avoid interference, usually the transmission between different RS will occupy different radio resources (e.g. subchannels for OFDM system). In 802.16e, there are 35 data subchannels. If 20 of them should be allocated to communication among the 20 RSs, the system efficiency will be heavily decreased. In the power line monitoring scenario, bi-directional antenna system will be used to enhance coverage along the power line. Thus, ways to maximize spatial and spectrum reuse will be helpful to increase system efficiency here.

- **Accurate synchronization:** 802.16j provides a good reference for TDD relay system design. Its proposed non-transparent mode is a classical design for the system

with more than 2 hops. But it will highly rely on the accurate synchronization on both UL and DL for all the access zone (BS/RS to MS) and relay zone (BS/RS to RS). Any error will generate accumulated impact for the system.

Besides these challenges, we should also pay attention to the selection of centralized/distributed scheduling mode, radio resource management, and channel quality information management, etc.

The RF part and antenna design are also main challenges for the monitoring system due to the long distance communication with limited solar power supply. We are inclined to obtain the qualified RRU through OEM. Hence, we will not discuss it further in this paper.

## 5 Conclusion and Future Work

In this paper, we propose a hybrid wireless monitoring system for power transmission system in China Smart Grid. The solution integrates the WiFi and WiMAX technology to support the two ways broadband communications. The system design is introduced in this paper and the key issues are addressed in this paper. And we also discuss the challenges for system implementation from spectrum optimization, low power, and TDD relay aspects. Through our work, we wish to improve the progress of China Smart Grid. Currently, we are working on the topics related to spectrum optimization and low power design.

## References

1. De Blasio, R.: Standards for the smart grid. In: Proceedings of IEEE Energy 2030 (2008)
2. Sarafi, A.M., Tsiropoulos, G.I., Cottis, P.G.: Hybrid wireless-broadband over power lines: A promising broadband solution in rural areas. *IEEE Communications Magazine*, 140–147 (November 2009)
3. Chen, T.M.: Smart grids, smart cities need better networks. *IEEE Network* (March - April 2010)
4. Lu, J., Xie, D., Ai, Q.: Research on smart grid in china. In: Proceedings of IEEE T&D Asia 2009 (2009)
5. Guo, J., Feng, Y., Ma, Y., Liu, G.: The future electric power communication network of strong and smart grid in china. In: 16th International Conference on Electrical Engineering, Busan Korea, July 11-14 (2010)
6. Sood, V.K., Fischer, D., Eklund, J.M., Brown, T.: Developing a communication infrastructure for the smart grid. In: Proceedings of IEEE Electrical Power & Energy Conference (2009)
7. Sun, F., Lei, M., Yang, C.: Establishing smart power grid and innovating management methods a new thought of the development of electric in china. IBM Corp. White Paper
8. Kahle, J.A., Day, M.N., Hofstee, H.P., Johns, C.R., Maeurer, T.R., Shippy, D.: Introduction to the cell multiprocessor. *IBM Journal of Research and Development* 49(4/5) (April 2005)
9. Wang, Q., Fan, D., Lin, Y., Chen, J., Zhu, Z.: Design of bs transceiver for ieee 802.16e ofdma mode. In: Proceedings of ICASSP 2008 (April 2008)

# Forgery Attacks of a Threshold Signature Scheme with Fault-Tolerance

Yingda Yu, Chuanda Qi\*, Junjie He, and Yuefen Chen

College of Mathematics and Information Science, Xinyang Normal University, Xinyang, China  
yuyd57@163.com, qichuanda@sina.com,  
hejj99@163.com, yuefenchen@126.com

**Abstract.** The threshold signature scheme with fault-tolerance is analyzed, and it is pointed out that the dishonest inside participants can forge the signatures of the authorized subset. An improved threshold signature scheme with fault-tolerance is proposed, and it is shown that the new scheme can resist insider forgery attacks effectively by analyzing the security of the improved scheme.

**Keywords:** threshold signature scheme, fault-tolerant, forgery attacks, discrete logarithm.

## 1 Introduction

Since the Public-Key Cryptography had been introduced by Diffie and Hellman, digital signatures are widely used.

In many cases, however, a file is sometimes necessary to sign by more than one person, and sometimes a signer need to designate an agent to sig. In 1979, Shamir[1] and Blakley[2] independently proposed a Secret Sharing Scheme, and made the  $(t, n)$  threshold sharing system. In 1992, Desmedt and Frankel[3] proposed the first  $(t, n)$  threshold digital signature scheme based on the RSA system. Subsequently, Harn[4] also proposed a threshold signature scheme, which is based on Shamir's secret sharing scheme and improved ELGamal signature scheme.

In threshold signature scheme, the sub-key holders need to make their own sub-signatures. But in the multi-party computing environment, it is difficult that to ensure the honesty of each participant. In paper [5], X.L. Zhang presented a threshold Signature Scheme with Fault-Tolerance via the principle of multi-party computation based on paper [6], which makes the threshold signature secure and effective in case of a few malicious behaviors.

In this paper, we analyzed the signature scheme which Zhang proposed in the paper [5], and pointed out that it is easy for the inside participants to forge signature, thus explained that Zhang's scheme is insecure, and also unable to be used in the actual. The structure of this paper as follows: in Section 2, we introduced the threshold signature scheme with fault-tolerance which was proposed by X.L. Zhang; in Section 3, we analyzed Zhang's scheme, and gave the method of forging an

---

\* Corresponding author.



authorized subset’s signature by inside participants; in Section 4 we improved Zhang’s scheme, and pointed out that the improved scheme can resist forgery attacks by the inside participants effectively after security analyzing.

## 2 The Threshold Signature Scheme with Fault-Tolerant

In [5], Zhang Xing-lan presented a threshold signature scheme with fault tolerance, via the principle of multi-party computation based on [6]. The signature process is as follows.

Assume  $P = \{p_1, p_2, \dots, p_n\}$  is a set of  $n$  participants, and  $U$  is a collection of subsets of  $P$ .  $U$  is called an access structure and the subsets in  $U$  are called authorized subsets, if the subsets in  $U$  are those subsets of participants that should be able to compute the secret. Suppose a  $(t, n)$  threshold scheme is to share a secret  $k$  among  $n$  participants, for each authorized subset  $A \in U$ , let  $A = \{p_1, p_2, \dots, p_l\}$ ,  $s_i$  is the shared secret value ( sub-secret ) about secret  $k$  of participant  $p_i$ ,  $x_i$  is the public parameter of  $p_i$  ( $i = 1, 2, \dots, l$ ).

### 2.1 Fault-Tolerant Protocol

Suppose there are at least  $t$  ( $l/2 < t$ ) honest participants in  $A$ , each participant  $p_i \in A$  runs the following secret sharing scheme once:

- (1)  $p_i$  randomly selects a degree  $t-1$  ( $t < l$ ) polynomial

$$f_i(x) = \sum_{k=0}^{t-1} a_{i,k} x^k = a_{i,t-1} x^{t-1} + a_{i,t-2} x^{t-2} + \dots + a_{i,1} x + a_{i,0}$$

over  $GF(q)$ , where  $f_i(0) = s_i$ .

Then, calculates  $A_{i,k} = g^{a_{i,k}} \bmod p$  ( $0 \leq k \leq t-1$ ), and send  $A_{i,k}$  to all the other participants in  $A$ .

- (2) After receiving  $\{A_{j,k}, j = 1, \dots, l; k = 0, \dots, t-1\}$ ,  $p_i$  computes  $\delta_{i,j} = f_i(x_j) \bmod q$  ( $j = 1, \dots, l, j \neq i$ ), and send  $\delta_{i,j}$  to  $p_j$  via a secure channel.

- (3) There is at least  $t$  honest parties in this sharing scheme, so any  $t$  honest parties can recover the polynomial  $f_j(x)$  by using interpolation with public point set  $\{(x_k, \delta_{j,i_k}) \mid k = 1, \dots, t\}$ , and then can calculate the secret  $f_j(0) = s_j$  of  $p_j$ .

Note: whether to set up  $g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_k} \bmod p$ , can be used to test whether the participant provides the correct sub-secret in the secret recovery process, in order to test the honesty of the participant.

### 2.2 System Initialization

The trusted center Dealer determines and calculates the following public parameters:

(1) Choose a secure Hash function;

(2) Choose two large prime numbers  $p$  and  $q$ , where  $q$  is a large prime factor of  $p-1$ , such that the discrete logarithm problem on  $Z_p^*$  is difficult. Generally,

$$2^{511} \leq p \leq 2^{512}, 2^{159} \leq q \leq 2^{160};$$

(3) Choose an element  $g$  of order  $q$  in  $Z_p^*$ ;

(4) Dealer calculates and publishes  $y = g^k \text{ mod } p$ . Participant  $p_i \in A$  calculates and publishes  $y_i = g^{s_i} \text{ mod } p$ . The secret  $k$  and the sub-secret  $s_i (i = 1, \dots, l)$  satisfy

$$k = \sum_{p_i \in A} c_i s_i,$$

where  $c_i \in GF(q)$  can be calculate using public parameters by any participant.

### 2.3 Threshold Signature Scheme with Fault-Tolerant

Running the following procedure, a threshold signature scheme with fault-tolerant can be generated effectively.

(1)  $p_i \in A$  selects a degree  $t-1 (t < l)$  polynomial

$$f_i(x) = \sum_{k=0}^{t-1} a_{i,k} x^k = a_{i,t-1} x^{t-1} + a_{i,t-2} x^{t-2} + \dots + a_{i,1} x + a_{i,0}$$

over  $GF(q)$  randomly, where  $f_i(0) = s_i$ . Calculates

$A_{i,k} = g^{a_{i,k}} \text{ mod } p (0 \leq k \leq t-1)$ , and sends  $A_{i,k}$  to all the other participants in  $A$ .

(2) After receiving  $\{A_{j,k}, j = 1, \dots, l; k = 0, \dots, t-1\}$ ,  $p_i$  computes

$$\delta_{i,j} = f_i(x_j) \text{ mod } q (j = 1, \dots, l, j \neq i),$$

and send  $\delta_{i,j}$  to  $p_j$  via a secure channel.

(3) Received the shared values of sub-secret  $s_j$ ,  $p_i$  checks

$$g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_i^k} \text{ mod } p.$$

If it is correct, we continue next step, otherwise start again.

(4) An authorized subset  $A$  threshold signs the message  $m$ :

Each participant  $p_i \in A$  runs the above threshold signature once. If there is a sub-signature of  $p_j \in A$  can't be verify, each participant  $p_i (i \neq j)$  publishes the shared values of sub-secret  $s_j$ , and checks the correctness of shared values by

$$g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_i^k} \text{ mod } p.$$

Because there are at least  $t$  honest parties, using their  $\delta_{j,i}$  checked by

$$g^{\delta_{j,j}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_j^k} \bmod p,$$

the polynomial  $f_j(x)$  can be recovered by interpolation. Then we can calculate the  $p_j$ 's secret  $f_j(0) = s_j$ , and replace  $p_j$  to sign again with the recovered secret.

(5) Each participant  $p_i \in A$  calculates the sub-signature of message  $m$  as follows: randomly selects an integer  $b_i \in [0, q-1]$ , computes and publishes  $r_i = g^{b_i} \bmod p$ , and then calculates

$$\begin{aligned} \delta_i &= H(m)b_i + (c_i + 1)s_i \bmod q, \\ \text{Sig}_i(m) &= (r_i, \delta_i). \end{aligned}$$

Checks the equation  $g^{\delta_i} = r_i^{H(m)} y_i^{c_i+1}$ . If it is correct, the sub-signature is valid, otherwise, it is invalid.

(6) Calculates

$$\begin{aligned} R &= \prod_{p_i \in A} r_i \bmod p, \\ S &= \sum_{p_i \in A} \delta_i \bmod q. \end{aligned}$$

The threshold signature is  $(R, S)$ .

(7) Authorized subset  $A$  computes

$$Y_A = \prod_{p_i \in A} y_i \bmod p.$$

Checks the equation  $g^S = R^{H(m)} Y_A y \bmod p$ . If it is correct, the threshold signature is valid, otherwise, it is invalid.

### 3 Insider Attack against the Threshold Signature Scheme with Fault-Tolerant

By analysis, we find that the threshold signature scheme with fault-tolerant which is introduced in the previous section is insecure, and any inside participant can easily forge the signature of an authorized subset. The method of forgery is as follows:

The inside participant let

$$R = g(Y_A^{-1} y^{-1})^{H(m)^{-1}} \bmod p.$$

In order to ensure the equation

$$g^S = R^{H(m)} Y_A y \bmod p,$$

the inside attacker creates the following equation

$$\begin{aligned}
 g^S &= (g(Y_A^{-1}y^{-1})^{H(m)^{-1}})^{H(m)} Y_A y \bmod p \\
 &= g^{H(m)} (Y_A^{-1}y^{-1}) Y_A y \bmod p \\
 &\equiv g^{H(m)} \bmod p.
 \end{aligned}$$

Then the attacker computes

$$S \equiv H(m) \bmod p .$$

The tuple  $(R,S)$  is a valid threshold signature of message  $m$ .  
In fact,

$$\begin{aligned}
 &R^{H(m)} Y_A y \bmod p \\
 &= g^{H(m)} ((Y_A^{-1}y^{-1})^{H(m)^{-1}})^{H(m)} Y_A y \bmod p \\
 &= g^{H(m)} ((Y_A^{-1}y^{-1}) Y_A y \bmod p \\
 &= g^{H(m)} \bmod p \\
 &\equiv g^S \bmod p.
 \end{aligned}$$

So

$$g^S = R^{H(m)} Y_A y \bmod p ,$$

that is the tuple  $(R,S)$  satisfy the verification equation, and it is a valid threshold signature of message  $m$ .

Any dishonest inside participant can forge a valid threshold signature of message  $m$  by structuring  $R = g(Y_A^{-1}y^{-1})^{H(m)^{-1}} \bmod p$ . Thus, Zhang's threshold signature scheme with fault-tolerant is insecure. Designer thought that the scheme can resist the attack by a few dishonest inside participants. And also the computation is reduced, because the participants are able to complete the sub-signature independently, and only calculate  $R = \prod_{p_i \in A} r_i \bmod p$  when generating the threshold signature after the signature of each participant is validated. But from the above analysis, we can see that the authorized subset's threshold signature of message  $m$  can be forged even if only one dishonest inside participant. In other words, Zhang's threshold signature scheme does not have the unforgeability.

#### 4 Improvement of the Threshold Signature Scheme with Fault-Tolerant

In order to resist the forgery attack of inside participants, we have improved the above threshold signature scheme with fault-tolerance. The improved scheme consists of the following five phases: system initialization phase, the key generation phase, the sub-signature generation and verification phase, the threshold signature with fault-tolerance generation phase and the signature verification phase.

### 4.1 System Initialization

The trusted center Dealer chooses the following public parameters:

- (1) Choose two large prime numbers  $p$  and  $q$ , where  $q$  is a large prime factor of  $p-1$ , such that the discrete logarithm problem on  $Z_p^*$  is difficult. Generally,  $2^{1023} \leq p \leq 2^{1024}$ ,  $2^{159} \leq q \leq 2^{160}$ ;
- (2) Choose an element  $g$  of order  $q$  in  $Z_p^*$ ;
- (3) Choose a secure Hash function;

### 4.2 Key Generation Phase

(1) First, each member  $p_i (i = 1, 2, \dots, l)$  randomly selects  $x_i \in [1, q-1]$  as his unique identification number, and publishes  $x_i$ . Then according to predetermined threshold  $t (t < l)$ , each member  $p_i$  randomly selects a degree  $t-1 (t < l)$  polynomial  $f_i(x) = \sum_{k=0}^{t-1} a_{i,k} x^k$  on  $GF(q)$ , where  $f_i(0) = s_i$  is the private key of  $p_i$ , and keeps secret. Calculates  $A_{i,k} = g^{a_{i,k}} \bmod p (0 \leq k \leq t-1)$ , and send  $A_{i,k}$  to all other participants in the authorized subset  $A$ .

(2)  $p_i$  computes  $\delta_{i,j} = f_i(x_j) \bmod q (j = 1, \dots, l, j \neq i)$ , and send  $\delta_{i,j}$  to  $p_j$  secretly.

(3) After receiving  $\{A_{j,k}, j = 1, \dots, l; k = 0, \dots, t-1\}$ ,  $p_i$  checks

$$g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_i^k} \bmod p.$$

The above equation used to test whether the participant provided the correct sub-secret in the secret recovery process, in order to test the honesty of the participant.

(4) Dealer computes and publishes  $y = g^k \bmod p$ . Each participant  $p_i \in A$  computes and publishes  $y_i = g^{s_i} \bmod p$ . It satisfies the relationship  $k = \sum_{p_i \in A} c_i s_i$  between the key  $k$  and  $s_i (i = 1, \dots, l)$ , where  $c_i \in GF(q)$ .

### 4.3 Sub-signature Generation and Verification Phase

Each participant  $p_i \in A$  computes the sub-signature of message  $m$  as follows:

(1) Select an integer  $b_i \in [1, q-1]$ , and computes  $r_i = g^{b_i} \bmod p$  randomly. Keeps  $b_i$  secret and publishes  $r_i$ .

(2) After receiving all  $r_j$ ,  $p_i$  computes

$$R = \prod_{p_i \in A} r_i \bmod p.$$

(3)  $p_i \in A$  computes the sub-signature  $\delta_i$  by using his private key  $s_i$ , the secret coefficient  $c_i$  and the random number  $b_i$ :

$$\delta_i = H(R, m)b_i + (c_i + 1)s_i \text{ mod } q. \tag{1}$$

$p_i$  sends the sub-signature  $(r_i, \delta_i)$  to the authorized subset  $A$ .

(4) The authorized subset  $A$  verifies

$$g^{\delta_i} = r_i^{H(R, m)} y_i^{c_i + 1} \text{ mod } p.$$

If it is correct, the sub-signature is valid, otherwise, it is invalid.

#### 4.4 Threshold Signature with Fault-Tolerance Generation Phase

(1) Each participant  $p_i (i = 1, 2, \dots, l)$  in  $A$  runs the above threshold signature once. If there is a sub-signature of  $p_j \in A$  can't be verified, each participant  $p_i \in A (i \neq j)$  publishes the shared values of sub-secret  $s_j$ , and checks the correctness of shared values by  $g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_i^k} \text{ mod } p$ . Because there are at least  $t$  honest participants, the sub-secret  $s_j$  can be recovered by using any  $t$   $\delta_{j,i}$ s which have checked by  $g^{\delta_{j,i}} = \prod_{k=0}^{t-1} (A_{j,k})^{x_i^k} \text{ mod } p$  with interpolation, and replace  $p_j$  to sign again with the recovered secret.

(2) After receiving the sub-signatures, the authorized subset  $A$  computes

$$S = \sum_{p_i \in A} \delta_i \text{ mod } q.$$

Finally, the threshold signature is  $(R, S)$ .

#### 4.5 Signature Verification Phase

The signature verifier calculates

$$Y_A = \prod_{p_i \in A} y_i \text{ mod } p,$$

and checks the equation

$$g^S = R^{H(R, m)} Y_A y \text{ mod } p. \tag{2}$$

If it is correct, the threshold signature is valid, otherwise, it is invalid.

### 5 Security Analysis of the Improved Signature Scheme

The security analysis of the improved signature scheme is as follows.

(1) The attacker can not get the secret  $k$  from the public  $y$ , because he will face the discrete logarithm problem.

(2) It is impossible that the insider attacker  $D$  should try to impersonate a legal participant  $p_i$  and forge his sub-signature.

Suppose the insider attacker  $D$  try to forge the sub-signature of a legal participant  $p_i$ .  $D$  randomly select  $u_i$ , let  $r_i' = g^{u_i} y^{v_i} \bmod p$ , where  $v_i \in Z_q (i=1, \dots, l)$  is undetermined. By the sub-signature formula (1), we can see  $H(R, m)$  in  $\delta_i$  is not only related to the message  $m$ , but also related to the personal information and private key of the participant. If  $D$  can forge a sub-signature  $(r_i', \delta_i')$ , it is show that  $D$  can get the sub-key  $s_i$  of the participant  $p_i$ . But  $D$  can not get the value of  $b_i$  because of the difficulty of Discrete Logarithm Problem on  $Z_p^*$ . It is impossible that  $D$  can get  $s_i$  by solving equation

$$\delta_i = H(R, m)b_i + (c_i + 1)s_i \bmod q.$$

And then the insider attacker  $D$  can not submit complete sub-signature. So the attacker  $D$  can not impersonate a legal participant  $p_i$  to forge a valid sub-signature.

(3) It is feasible that the insider attacker  $D$  should try to impersonate an authorized subset  $A$  to forge a valid threshold signature with fault-tolerance of message  $m'$ .

Suppose the insider attacker  $D$  impersonates the authorized subset  $A$  and tries to forge a valid threshold signature with fault-tolerance of message  $m'$ . By the signature synthesis formula (2), we can see  $H(R, m)$  is not only related to the message  $m$ , but also related to the personal information and private keys of the participants. And then the attacker  $D$  can not submit complete group signature. In addition, If  $D$  wants to work out forged signature  $S'$ , he will face the discrete logarithm problem. Therefore, the attacker  $D$  can not impersonate authorized subset  $A$  to forge a valid threshold group signature.

## 6 Conclusion

Threshold signature scheme with fault-tolerance have special applications in electronic commerce and electronic government. As any dishonest inside participant can forge the signature of the authorized subset, X.L. Zhang's threshold signature scheme with fault-tolerance is insecure, In this paper, we proposed an improved signature scheme with fault-tolerance. The improved scheme can resist insider attacks effectively.

**Acknowledgments.** This work was supported by the Natural Science Foundation of Henan Province of China under Grant No. 102102210242.

Chuanda Qi (corresponding author): Phn: +86-376-6391030, E-mail: qichuanda@sina.com.

## References

1. Shamir, A.: How to share a secret. *Communications of the ACM* 22(11), 612–613 (1979)
2. Blakley, G.R.: Safeguarding cryptographic key. In: *Proc AFIPS 1979, National Computer Conference*, pp. 313–317. AFIPS press, New Jersey (1979)
3. Desmedt, Y., Frankel, Y.: Shared Generation of Authenticators and Signatures. In: Feigenbaum, J. (ed.) *CRYPTO 1991. LNCS*, vol. 576, pp. 457–469. Springer, Heidelberg (1992)
4. Harn, L.: (t,n) Threshold signature and digital multi-signature. In: *Workshop on Cryptography & Data Security Proceedings, R O C*, pp. 61–73 (1993)
5. Zhang, X.-L.: A Threshold Signature Scheme with Fault-Tolerance. *Journal of the Graduate School of the Chinese Academy of Science* 21(3), 398–400 (2004)
6. Xu, C.-X., Dong, Q.-K., Xiao, G.-Z.: A Vector-Space Secret Sharing Multisignature Scheme. *Acta Electronica Sinica* 31(1), 48–50 (2003)



# Performance Assessment and Detection of Instant Messaging Software Based on the AHP&LCA Model

ZongRen Liu and LeJun Chi

School of Computer Science and Technology,  
Harbin Institute of Technology at WeiHai, WeiHai, China  
lzongren@gmail.com, qdclj@163.com

**Abstract.** Due to the rapid development of Internet, new forms of Internet based services appear, whose desktop clients are similar to traditional instant messaging software (IMS). In addition, because the IMS's frequent updates and encryption of communication protocols, the traditional way of detecting IMS is not suitable anymore. So we promote a model combined with Analytic Hierarchy Process (AHP) and Life Cycle Assessment (LCA), to complete the performance assessment and detection of new forms of IMS and traditional ones, which based on advanced packet capture and analysis using Distribute Computing. The experimental results show that the performance of this method is effective and efficient. The most significant contribution of the AHP&LCA model in IMS is its flexibility which suits various needs. This paper provides a new thought of detecting of IMS.

**Keywords:** instant messaging, performance assessment, detection, AHP, LCA.

## 1 Introduction

Although the research of detecting of Instant Messaging Software (IMS) starts since its birth, there come some new problems waiting to be solved. Firstly, with the rapid development of Internet, many new forms of Internet based services such as Social Networking Service (SNS) and Microblog come out. These services have also promoted desktop clients to attract more users and become very popular. These clients software can also complete the instant messaging function as the traditional IMS do but differ from them. Secondly, the traditional way to detect IMS is mainly focused on the analysis the transcript [1]. It is often not effective and efficient because some of the IMS's communication protocols are encrypted and update frequently. Thus, it's difficult to subtract information from the packet captured constantly and correctly.

In order to solve the problems presented above, we have promoted new ideas to complete the work of detecting of IMS. Briefly, we come up with a model combined with Analytic Hierarchy Process (AHP) and Life Cycle Assessment (LCA) to evaluate the performances of IMS. And the performances of IMS are based on the analysis and detecting of IMS which the promoted packet capture and analysis of IMS's communication protocols are required. By taking advantages of the mathematical model, we could not only detect the state of IMS, but also obtain the

performances of certain IMS. As to the desktop client software of the new Internet based services, we define them as the Instant Messaging Like Software (IMLS). The IMLS have also been considered in our work. To evaluate our method in practice, we test it in real network environment. And the experimental results show that our approach can work prospectively.

## 2 Nomenclature

The nomenclatures used in this paper and the definitions of them are listed in Tab.1.

**Table 1.** Nomenclatures

Nomenclature	Definition
AHP	Analytic Hierarchy Process
LCA	Life Cycle Assessment
G	Goal (AHP)
C	Criterion (AHP)
A	Alternative (AHP)
S	Solution (AHP)

## 3 Related Work

In order to show the universality, we take different types of IMS and IMLS into consideration. Each of them represents one type of applications. Tab.2 shows the titles of the software in the study.

**Table 2.** Details of software studied

Software	Description
Google Talk (GTalk)	Traditional IMS, stands for IMS
MSN	Traditional IMS, stands for IMS
Sina Weibo	New IMS, Microblog, stands for IMLS
RenRen Desktop	New IMS, SNS, stands for IMLS

The AHP&LCA model requires to classify the actions of the IMS. According to numerous popular IMS, the actions can be divided into five types blew [2]. Tab.3 shows the five types.

In addition, not all the software in Tab.2 contains the whole five action types.

**Table 3.** Action types of IMS

Action Types	Description
Transcript	Text based conversations
Audio	Audio chat
Video	Video chat
File transfer	User transfers files to another
Control information	Information exchanged with IMS Server

### 3.1 Analysis of Communication Protocols of IMS

As to the four software listed in Tab.2, GTalk, Sina Weibo and RenRen desktop are based on the Extensible Messaging and Presence Protocol (XMPP), which is widely used and has been studied well [4, 5]. The remaining one, MSN, is using its own private protocol, known as the Microsoft Notification Protocol (MSNP).

According to our approach, the XMPP based IMS are working similarly. They developed the model of XMPP and fill their data in packet with the format of Extensible Markup Language (XML). The data is mainly contain in the three XMP elements, which are `<message/>`, `<presence/>` and `<ip/>`. For example, the child element `<body/>` of `<message/>` often contains the transcripts if the *type* attribute of `<message/>` is set as *default*. In addition, the communication port of them is always constant, such as port 5222.

To MSN, the latest MSNP is version 19. Though the MSNP is private, we concluded some certain features. For example, by recognize the port vector, the *MSN* sign and the two */r/n* sign in packets, we can easily subtract the transcript text in UTF-8 format.

So based on the work of analysis of communication protocols of IMS, we can capture the certain packets of the five actions types in Tab.3. Thus, the performance evaluation and detecting of such IMS are practical.

### 3.2 Packet Capture and Analysis Environment

Briefly, we developed the packet capture component by using the WinPcap library with the results presented in Related Work above. And to increase the efficiency of packet capturing and analysis, we introduce the Distributed Computing using Microsoft Distributed Component Object Model (DCOM) [3].

Fig.4 shows the environment of the packet capture component.

The whole packet capture and analysis environment is constructed in the local area network (LAN). The *Monitored Host* is using IMS. The *Packet Capture Host* mainly runs the component of packet capture. It captures all the IMS packets of LAN, then restores them and analyses them. If the load is high enough, the Distributed Computing program starts up, transferring certain packets to the *Distributed Computing Host*, which completes the analysis work until the working lower down, then stops. The data of the whole system is stored in the database presented in the figure.

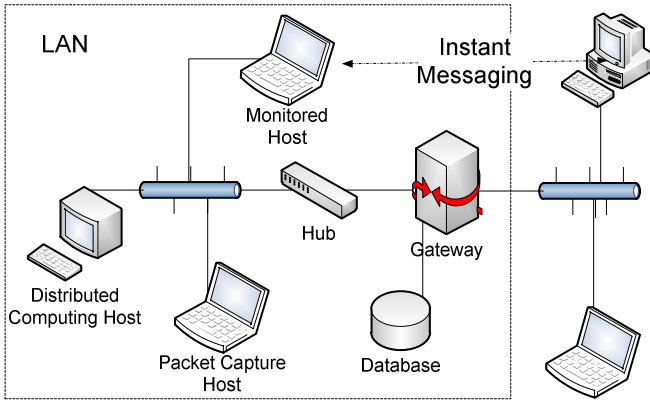


Fig. 4. Packet capture & analysis environment

## 4 Model of AHP&LCA

### 4.1 Introductions

The traditional LCA model works straightforward, and mainly contains four steps blew:

- Goal & Scope Definition
- Inventory Analysis of Result
- Impact Assessment
- Interpretation

To make the model suit the situation of IMS detecting, we improve the method of using AHP in LCA. Fig.5 shows the promoted model.

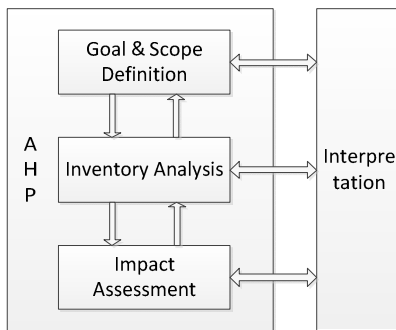


Fig. 5. AHP&LCA model

### 4.2 Modeling: Goal Definition

Using AHP&LCA model firstly come the definition of LCA’s goal the building hierarchy of AHP. There are following steps.

- LCA Goal Definition

Based on the needs of detecting IMS, we define the goal as the performance of using IMS by certain host. Here we mainly consider whether the host is working and not just killing time.

- Build hierarchy of AHP

Fig.6 shows the hierarchy of AHP. And the nomenclatures used here have been defined in Tab.1.

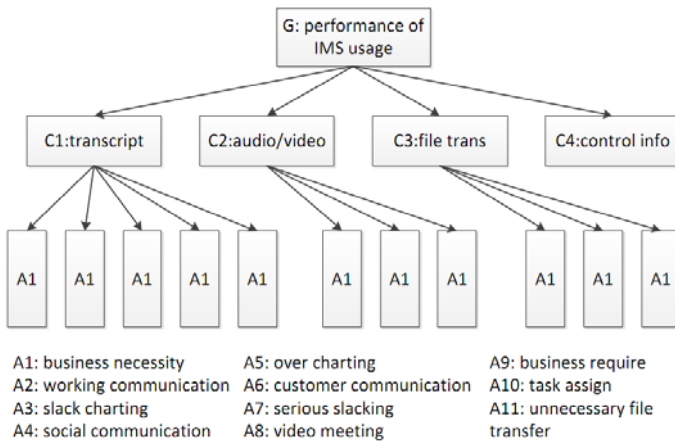


Fig. 6. Hierarchy of AHP

In Fig.6, there are four criteria and eleven alternatives. The criteria here are based on the classification of IMS’s action types in Tab.3. And the alternatives are proposed using Expert Investigation Method.

### 4.3 Modeling: Inventory Analysis with AHP

To complete the inventory analysis, we do the following steps to determine the weight.

- Build the judgment matrix

We use  $a_{ij}$  to estimate the importance which  $a_i$  compares to  $a_j$  . Thus the judgment matrix can be expressed as

$$\alpha = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = (a_{ij})_{n \times n}$$

In which,  $a_{ij} > 0, a_{ij} = 1/a_{ji}, a_{ii} = 1$ .

With the data Expert Investigation Method provided, we can forward matrixes below:

$$M_1 = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 1/3 & 1 & 2 & 3 \\ 1/5 & 1/2 & 1 & 2 \\ 1/7 & 1/3 & 1/2 & 1 \end{pmatrix} \quad M_3 = \begin{pmatrix} 1 & 1/5 & 3 \\ 5 & 1 & 9 \\ 1/3 & 1/9 & 1 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 1 & 4 & 2 & 8 & 6 \\ 1/4 & 1 & 1/2 & 2 & 2 \\ 1/2 & 2 & 1 & 4 & 3 \\ 1/8 & 1/2 & 1/4 & 1 & 1 \\ 1/6 & 1/2 & 1/3 & 1 & 1 \end{pmatrix} \quad M_4 = \begin{pmatrix} 1 & 1 \\ 7 & 1 \end{pmatrix}$$

- Normalization of columns of matrixes

$$\bar{a}_{ij} = a_{ij} / \sum_{k=1}^n a_{kj} \quad (i, j = 1, 2, \dots, n)$$

- Calculating the largest eigenvalue & the eigenvector of the judgment matrix

$$\bar{w}_i = \sum_{j=1}^n \bar{a}_{ij} \quad w_i = \bar{w}_i / \sum_{i=1}^n \bar{w}_i$$

$$\lambda_{\max} = \frac{1}{n} \sum_i \frac{(AW)_i}{w_i}$$

After calculating, we reach the eigenvector as below:

$$w_{total} = (0.4880, 0.1298, 0.2440, 0.0650, 0.0733)^T$$

- Consistency check

If  $CR = \frac{CI}{RI} < 0.1$ , the consistence of the judgement matrix can be accepted.

Unless, the matrix should be correct.

$$CR = \frac{\lambda_{\max} - n}{n - 1}$$

**Table 7.** Mean random consistency index of RI

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

#### 4.4 Modeling: Impact Assessment

According to the consistency check, the result can be accepted. To be cleaner and more intuitionistic, we form Tab.8 below to show the weight of each factor.

These factors below can be used as weight to calculate the performance of IMS.

#### 4.5 Strengths and Weakness

The processes of AHP&LCA model are presented above. We can clearly know the strengths and weakness of the model.

The most obvious character of the AHP&LCA model is its flexibility. The basic judgment matrix can be defined according to different needs. For example, if the user of the AHP&LCA, totally disagrees with using IMS's video chatting, he or she could probably set the judgment value of video obvious enough to show the concern. But to be fairly, the judgment matrix is better to be set by Expert Investigation Method, if it is applied in a company, the managers and administrative stuffs could serve as the experts. The values of the matrix we use are the consideration values.

**Table 8.** Weight of each factor

	C1	C2	C3	C4	Weight
	0.5828	0.2449	0.1245	0.0730	
<b>A1</b>	0.4880	-	-	-	0.2844
<b>A2</b>	0.1298	-	-	-	0.0756
<b>A3</b>	0.2440	-	-	-	0.1422
<b>A4</b>	0.0650	-	-	-	0.0379
<b>A5</b>	0.0733	-	-	-	0.0423
<b>A6</b>	-	0.1804	-	-	0.0442
<b>A7</b>	-	0.7481	-	-	0.1832
<b>A8</b>	-	0.0715	-	-	0.0175
<b>A9</b>	-	-	0.5000	-	0.0623
<b>A10</b>	-	-	0.5000	-	0.0623
<b>A11</b>	-	-	0.5327	-	0.0602

## 5 Performance Assessment and Detecting of IMS

### 5.1 Design of Packet Capture and Analysis

The environment of packet capture and analysis is shown in Fig.4. The main procedures of this component are designed as Fig.9 shows.

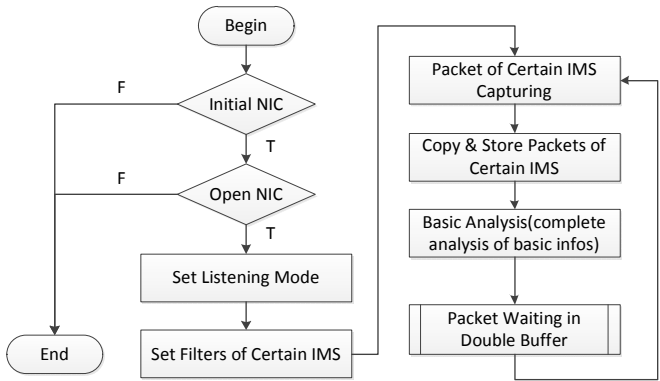


Fig. 9. Design of packet capture & analysis

In Fig.9, the design of packet capture and analysis is presented. The NIC in the figure refers to the network interface card (NIC). Firstly, the NIC needs to be initiated. Thus it can be opened to work. In order to capture all the packets through LAN, the listening mode should be set as promiscuous model. The filter means the characters of packets, according to the approach in Section 3.1, the filter can be set as the four certain IMS. Then the program is working to capture all the packets of the certain IMS. Then the basic analyser subtracts the basic information of the packets. After that, the packets are collected in the double buffer area waiting to be dealt by followed programs.

In addition, if the load of analysis is too high, the Districted Computing will start to lower the load. But it is not the key issues, so it's not presented in Fig.9.

5.2 Performance Assessment

Because of the efforts made in packet capture, the performance evaluation is not complicate. Fig.10 describes the procedures of the performance assessment.

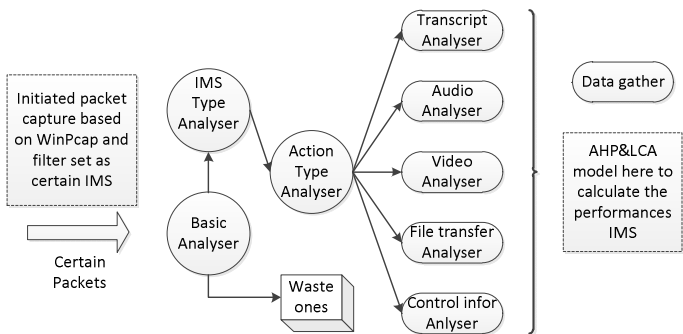


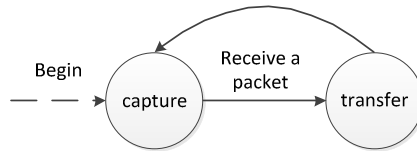
Fig. 10. Main procedures of performance evaluation



Certain packets of IMS would be firstly transported to *Basic Analyser*, which subtract basic information such as port, IP and drop waste ones. Then the key parts of the packets are delivered to *IMS Type Analyser* and the IMS type would be recognized. Thus, the main parts of packets could be decomposed and analysed by the *Action Type Analyser*, which involves five smaller analyser known as *Transcript Analyser*, *Audio Analyser*, *Video Analyser* and *Control Information Analyser*. After all the processes, the data would be gathered together, thus the model based on AHP&LCA can assess the performance of the IMS.

## 6 Experimental Results

The AHP&LCA model's practical applicability is mainly determined by its working performance. In order to complete the efficiency test, we design a network flooding like environment. The environment is very similar to Fig.4. The difference is the additional packet producer program, which mechanism is showed in Fig.11.



**Fig. 11.** Automation of performance test

With the program, we gathered the data below in Tab.12.

**Table 12.** Overall experimental data

Test Time(min)	Packet Number	All Packet Size(MB)	Average Speed(/second)
623	198,784,362	9861.12	5521

In order to complete the test without dependence of machine, we use the times the double buffer blocks to estimate. We test the size of the buffer from 1 to 1000, which its unit is 8 bytes. Fig.13 shows the result.

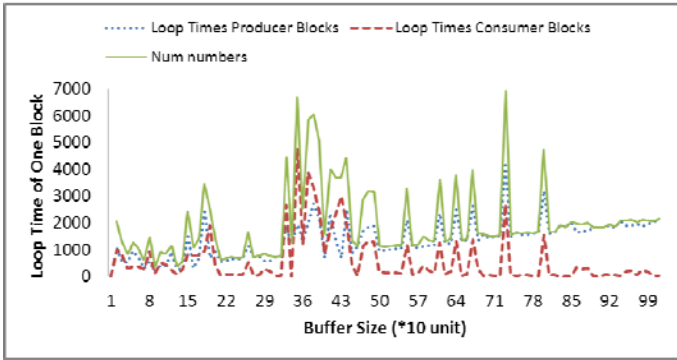


Fig. 13. Loop times of blocking

According to the test environment, the machine’s CPU main frequency is 2.75G. We assume that the packet sending frequency as  $G$ . And it is related with CPU main frequency, capturing time and the environment. So we can reach the formula below.

$$G = \lambda * f(g, t, s)$$

If the loop times when block happens is  $K$ , then the blocking time is:

$$T = G / K$$

In the worst condition, the blocking times would not exceed the buffer size  $B$ , so the dealing time is:

$$T_d = G / K * B$$

According to the result in Tab.12, we assume the speed of packet capture as  $M$ . We can reach the dealing time of those packets, then we can forward the host numbers this program can sustain:

$$1 / (T_d * M)$$

If we choose one of the best sizes of buffer as 20 units from Fig.13, we reach the result 36, which means it can afford 36 hosts in normal network load.

In addition, if the Distributed Computing Program is used, more hosts can be sustained.

**Acknowledgments.** The authors would like to thank the anonymous reviewers for their helpful comments for improving this paper.

Research and implementation of data acquisition and management system are based on RFID (Project of Natural Science Foundation of Shandong Province, serial NO. 2006ZRA10010).

## References

1. Zhang, H., Xu, G., Li, J., Wang, L.: Research on Detection of Instant Messaging Software. In: Qi, L. (ed.) ISIA 2010. CCIS, vol. 86, pp. 664–669. Springer, Heidelberg (2011)
2. Niinimäki, T., Lassenius, C.: Experiences of Instant Messaging in Global Software Development Projects: A multiple Case Study. In: Proceedings - 2008 3rd IEEE International Conference Global Software Engineering, ICGSE 2008, pp. 55–64 (2008)
3. Wang, F.-G., Tian, D., Yang, H.-S., Sun, R.: A Distributed Monitoring System For Working Status of Science Instruments. In: Proc. IEEE International Conf. on Automation and Logistics, pp. 1016–1021 (2008)
4. Internet Engineering Task Force, Extensible Messaging and Presence Protocol (XMPP); Core (2004), <http://www.ietf.org/rfc/rfc3920.txt>
5. Day, M., Rosenberg, J., Sugano, H.: A Model for Presence and Instant Messaging. RFC 2778 (International) (February 2000)

# Chinese Accents Identification with Modified MFCC

Qin Yan<sup>1</sup>, Zhengjuan Zhou<sup>1</sup>, and Shan Li<sup>2</sup>

<sup>1</sup> College of Computer and Information, Hohai University, Nanjing, China

<sup>2</sup> College of Economic and Management,

Nanjing University of Aeronautics and Astronautics, Nanjing, China

yanqin@ieee.org, jaylove\_2008@hotmail.com, lishan@nuaa.edu.cn

**Abstract.** It is well known that performance of Chinese speech recognition system fluctuates sharply with variation of accents. Hence it is feasible to identify accents before recognition. Based on our previous analysis [1], it is discovered that the first two formants are more sensitive to Chinese accents than others. A modified Mel-frequency cepstral coefficients (MFCC) algorithm is proposed by increasing the filter distribution in lower and middle frequency range to accommodate the sensitivity of Chinese accents. Comparing with the GMM system based on traditional MFCC, the error rate of the GMM system based on modified MFCC declines by 1.8%.

**Keywords:** formant, MFCC, accent identification, GMM.

## 1 Introduction

In speaker-independent recognition system, speaker variability, such as gender, accent, age, speaking rate, and phones realizations, is the difficulties in speech recognition task. It is shown in [2] that gender and accent are the two most important factors in speaker variability. Usually, gender-dependent model is used to deal with the gender variability problem. The accent variability related to the differences of phonemic systems, phonotactic distributions, lexical realizations of words, phonetic realizations and so on is a more complex and challenging question [3].

The varieties of Chinese cover a huge area containing about one billion people [4]. In general, recognition rate of a good robust standard Chinese recognition system declines by 20% to 50% in identifying accented Chinese [5]. A feasible method to solve this problem is to establish corresponding recognition model for each accent. The first step is to identify the speaker's accent before recognition, then select the corresponding model. It is effective to improve the accented Chinese recognition rate. [6] shows that recognition rate of accent-dependent system is higher than accent-independent system by about 30%. Therefore, improving accent identification rate has important practical significance.

Most of current accent identification systems are built based on the HMM framework. Although HMM is effective in classifying accents, its training procedure is time-consuming. Also, using HMM to model every phoneme or phoneme-class is computationally expensive. Furthermore, HMM training is a supervised fashion:

it needs phone transcriptions which are either manually labeled, or obtained from a speaker independent model, leading to degraded performance [6].

Mel-frequency cepstral coefficients (MFCC) are the popular parameters for accent identification. One of the key steps to get the parameters is dividing the linear frequency into triangular filter bank and output the bandwidth energy from each filter. The influence degrees of accent to formants are different[7]. Therefore, an algorithm to improve accent identification rate is to change the triangle filter density (such as increasing the triangular filter numbers of more sensitive frequency) of MFCC.

Studies employing modified MFCC to improve speech recognition rate have been carried out at home and abroad. [8] applies it to improve speaker identification rate in whispered speech since the locations of the formants and the auditory model in whispered speech are different from those in normal speech.[7] discovers that middle frequencies ( 1500~2500Hz) are sensitive to American English foreign accents and then proposes a modifying MFCC algorithm that increases middle filter density of MFCC. The speech recognition rate based on modified MFCC is higher than system based on traditional MFCC by 1.2%. Experimental corpus of [7] is only 20 independent words and four sentences. Overall, the studies of modifying MFCC are not enough. This paper employs rich accent Chinese corpus to analysis sensitive frequencies to Chinese accents, and then to prove the performance of the algorithm.

The paper is organized as follows. In section 2 the experimental corpus is briefed. Section 3 discusses the sensitive frequencies to Chinese accents. Section 4 proposes the modified-MFCC algorithm based on section 3. Section5 sets up the accent identification system based on GMM to prove the performance of the algorithm and section 6 concludes the paper.

**Table 1.** Corpus Statistics (Formant Analysis/Accent Identification)

	No. Speakers	Analysis/Training Corpus
CQ	10/10	1328/10 20 50
GZ	10/10	1490/10 20 50
SH	10/10	1490/10 20 50
XM	10/10	1478/10 20 50

**Table 2.** Formant Decreasing Ranking of Accents

	F1	F2	F3	F4
CQ/GZ	2	1	3	4
CQ/SH	1	2	3	4
CQ/XM	1	2	4	3
GZ/SH	2	1	3	4
GZ/XM	2	1	3	4
SH/XM	1	2	3	4

## 2 Experimental Corpus

The experimental corpus, only containing the female part, is selected from RASC863—A Chinese Speech Corpus with Four Regional Accents (namely Shanghai, Chongqing, Guangzhou and Xiamen) [9]. The average duration of formant analysis corpus is about 8s, whereas the duration of training corpus for accent identification is about 4s. The training corpus and test corpus are separate. Considering the performance of accent identification system based on GMM is quite well in the case of insufficient training corpus (3~5 sentences). In order to highlight the results, Chongqing dialect words, the duration of which is 1~2s, is chosen as the test corpus. Tab.1 is the experimental corpus of formant analysis and accent identification. Chinese with Chongqing, Guangzhou, Shanghai and Xiamen accent is briefed as CQ, GZ, SH and XM.

## 3 Sensitive Frequency Analysis

Estimation of first four formants (F1, F2, F3 and F4) of 34 finals in four accents have has been completed in our previous works[1]. In this section, the importance of each formant in conveying accent is assessed. The influences of four formants on accents are ranked according to a normalized distance measure as shown in equation (1)[3]. This formula is proposed for ranking the formants of an accent  $A$  with reference to the formants of another accent  $B$ , as

$$Rank(F_i) = \text{index}_{F_i}(\text{Sort}(\sum_{v=1}^V [\frac{F_{vi}^A - F_{vi}^B}{0.5(F_{vi}^A + F_{vi}^B)}]^2)) \quad i=1, \dots, N \quad (1)$$

where  $F_{vi}^A$  is the average gender-dependent  $i$ th formant of the final  $v$  from accent  $A$ ,  $V$  is the number of finals,  $N$  is the number of formants, and  $Rank(F_i)$  is obtained as the rank position of the  $i$ th formant after sorting the formants in (1) in the decreasing order of a normalized distance of the formants across the two accents.

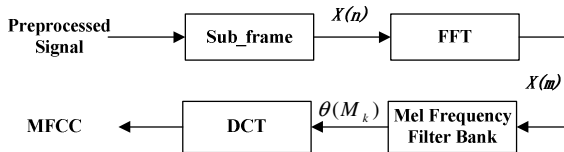


Fig. 1. MFCC Extraction Diagram

Equation (1) is used to obtain an estimate of the ranks of the contribution of formants to pairs of accents. The results, as shown in Tab.2, rank 1st and 2nd formants with the largest normalized differences across accents. The first two formants are more sensitive to accents than others. This may suggest that the first and second formants both are crucial formants for accents, different from [7] in which 2nd

formant are taken as the most accent sensitive formant. It could be explained by language differences. In the next section we propose a modified frequency scale for accent identification.

## 4 Modified MFCC Algorithm

### 4.1 Traditional MFCC

MFCC is the cepstral coefficient extracted from Mel frequency. It has good ability for recognition and noise immunity and has been proved to be one of the most successful characteristic descriptions in speech recognition tasks [10]. Fig.1 is the MFCC extraction diagram.

- 1) Original speech signal turns into linear frequency domain signal after pre-emphasis, framing, windowing, and FFT transforming.
- 2) Adding triangle bandpass filter in Mel frequency domain to get the filter bank and output the logarithmic energy of each filter.
- 3) Using DCT to remove the correlation of dimensional parameters and map them to the Low-dimensional space.

The sensitivity of human ear to normal speech is in vicinity of F1 (300~1000Hz). At the same time, F1 is also the energy concentration area. The energy of F1, F2, F3 and F4 decays in turn. The Mel scale filter bank is evenly placed in Mel domain as described in equation (2).

$$BW = 2 * Mel(f_s / 2) / (N + 1) \quad (2)$$

where BW is the triangular filter bandwidth placed in the Mel domain. The Mel function is used to transfer the linear frequency to Mel scale frequency. N is the number of triangular filters. In the view of linear frequency, Mel frequency band increases the weight of F1 where more filters are placed and makes MFCC more similar to human hearing system. Fig.2(a) is the Mel scale triangle filter bank with a sampling rate of 16KHz and 24 filters. Tab.3 displays the frequency of the mid-point of triangular filter bank.

### 4.2 Modified MFCC

As the analysis in section3, F1 and F2 are the two most important formants to accents. According to the statistics, F1 ranges from 300Hz to 1000Hz and F2 ranges from 1000Hz to 2600Hz. However in traditional MFCC, filters are equally distributed in Mel scale. Hence it is sensible to increase filter distribution in low and middle frequency range by assigning more filters.

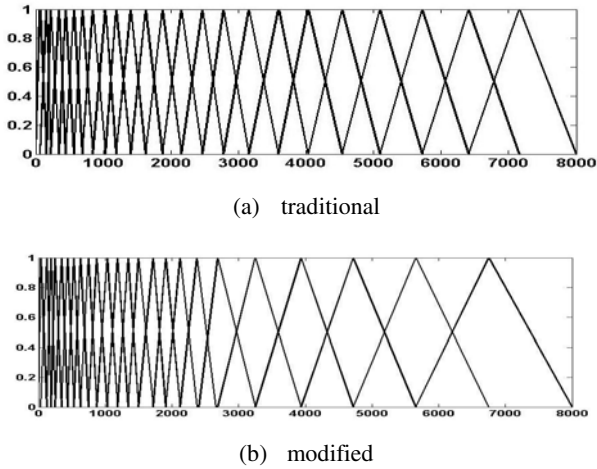


Fig. 2. Mel Scale Filter Bank

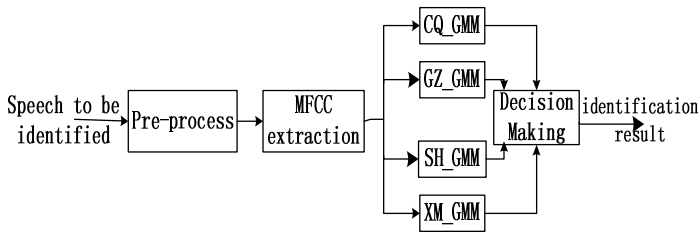


Fig. 3. Accent Identification Procedure

Section 3 indicates that it is feasible to evenly increase the number of filters under 2600Hz by reducing high frequency filter number appropriately since the human ear is not very sensitive to high frequency. Tab.3 shows that 15 filters under of 2600 Hz accounts for 62.5% of total number of filters in traditional MFCC. In order to increase filter density of the low and middle ranged frequency and have no loss of generality

(considering different sample rates), 75% (18 filters) of total filters are placed in the first 1/3 frequency band (0~2667Hz) with a sampling rate of 16KHz. The modifying formulae of triangular filters are as follows.

- 1)  $f < f_s / 6$  ,  $BW = 2 * Mel(fs / 6) / (N * 0.75 + 1)$  ;
- 2)  $f > f_s / 6$  ,  $BW = 2 * (Mel(fs / 2) - 2 * Mel(fs / 6)) / (N * 0.25 + 1)$  ;
- 3)  $f = f_s / 6$  ,  $BW = Mel(fs / 6) / (N * 0.75 + 1) + (Mel(fs / 2) - Mel(fs / 6)) / (N * 0.25 + 1)$

The Fig.2(b) shows the modified Mel scale filter bank. Tab.3 shows the mid-point frequency of the modified triangle filter bank.



## 5 Experiment and Analysis

### 5.1 System Setup

Accent identification, similar to other pattern recognition, includes three phases: feature extraction, model building and decision making.

In this paper, the main goal is to test the algorithm performance, the speech feature vectors consist of 36 features including 12 MFCCs and their first and second derivative features, not including the energy parameters. For each accent, a GMM is built, separately named CQ\_GMM, GZ\_GMM, SH\_GMM and XM\_GMM.

The output of accent identification is to find the accent model  $\lambda_i$  that the speech signal  $i^*$  belonged to and make the speech vector group  $X$  with the largest posterior probability  $P(\lambda_i / X)$ . According to Bayesian theory, the common method to obtain maximum value of  $P(\lambda_i / X)$  always employs  $P(X / \lambda_i)$  and the final decision rule is

$$i^* = \arg \max P(X / \lambda_i) \quad (3)$$

The accent identification procedure is shown in Fig 3. The first step is to training GMM for four accents, then to preprocess and extract the feature parameters of the speech to be identified and input them to each model. The last step is to choose the maximum value model as the accent belonging to the speech.

### 5.2 Experimental Analysis

Similar to common speech processing, the preprocessing of accent identification includes pre-emphasis (pre-emphasis coefficient chosen as 0.97), sub-frame (512 points per frame, frame shift is 160 points), windowing (Hamming window). In this section, MATLAB is employed as the development platform.

#### (1) Gaussian component N

Gaussian component N is very important to accent recognition rate. But oversize N may decline the identification rate of system. In this section, N is chosen as 1, 4, 8 respectively to compare the identification rate of systems based on traditional MFCC parameters and improved MFCC parameters.

#### (2) Scale of training corpus M

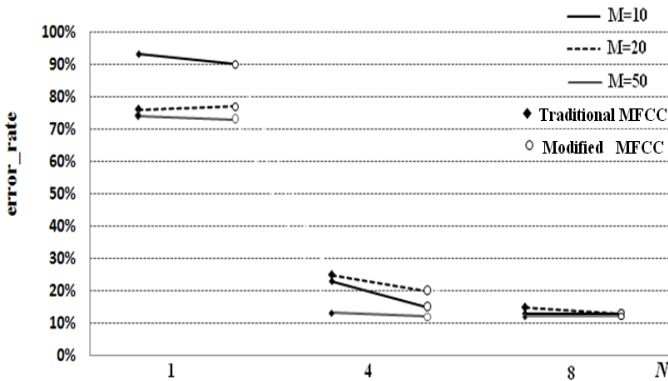
It is helpful for robustness of models by increasing the scale of training corpus M. But it will add to the computational burden of model. Therefore, selecting appropriate M is necessary. In this section, M is set as 10, 20, 50 to compare the performance of the modified system and the benchmark system respectively.

**Table 3.** The Mid-point Frequency of Triangle Filter Bank (traditional/modified)Hz

1	2	3	4	5	6
94/ 64	188/ 156	250/ 219	375/ 281	469/ 375	594/ 469
7	8	9	10	11	12
719/ 563	875/ 656	1063/ 781	1219/ 906	1438/ 1063	1656/ 1219
13	14	15	16	17	18
1906/ 1375	2188/ 1531	2500/ 1750	2813/ 1938	3188/ 2156	3625/ 2406
19	20	21	22	23	24
4063/ 2688	4563/ 3250	5125/ 3938	5750/ 4719	6438/ 5656	7188/ 6750

**Table 4.** Accent Identification Error Rate(traditional/modified)

	M=10	M=20	M=50
N=1	93%/90%	76%/77%	74%/73%
N=4	23%/15%	25%/20%	13%/12%
N=8	13%/13%	15%/13%	12%/12%

**Fig. 4.** Error Rate Comparison of Two Systems

The results of the experiment are shown in Tab.4. Fig.4 show the error rate changes more intuitively in the form of figure. In Fig.4, whether the MFCC extraction algorithm is modified or not, larger  $N$  and  $M$  both reduce the error rate of accent identification. When  $N = 1$ ,  $M = 10$ , the error rate is maximum and exceeds 90%. When  $N = 8$ ,  $M = 50$ , the error rate is smallest and stable at 12%. But when  $M = 50$ ,  $N$  increasing from 4 to 8, It seems increasing Gaussian components are of no benefit to the accent identification rate. Similarly when  $N = 8$ , increasing  $M$  brings no

advantage to the results. It could be concluded that N and M play complementary roles in the accent identification systems.

Fig.4 illustrates that the system based on modified MFCC has lower identification error rate than system based on traditional MFCC in most cases (except  $N=1$ ,  $M=20$ ). Overall the error rate of the system with modified MFCC declines by 1.8% compared to benchmark system with traditional MFCC on average.

## 6 Conclusion

This paper proposes a modified MFCC algorithm based on the fact that certain frequency range is more sensitive to Chinese accent than others. In order to increase filter density of the low and middle ranged frequency and have no loss of generality (considering different sample rates), 75% (18 filters) of total filters are placed in the first 1/3 frequency band (0~2667Hz) with a sampling rate of 16KHz. Comparing with traditional MFCC, the modified MFCC decreases accent identification error rate by 1.8%. In future work, more experiments will be made filter distribution in Mel scale for performance improvement.

**Acknowledgement.** We wish to thank the sponsorship of Hohai University Natural Science fund (2008428811), Doctoral fund of Education Ministry of China (200802941007), natural science fund of Jiangsu Province (No. 2010505411).

## References

1. Yan, Q., Zhou, Z., Xu, Z.: Acoustics Analysis of Differences across Formants of Four Chinese PUTONGHUA. In: International Conference on Multimedia Technology (ICMT), pp. 1–3 (2010)
2. Huang, C., Chen, T., Li, S., Chang, E., Zhou, J.L.: Analysis of Speaker Variability. In: Proc. Eurospeech 2001, vol. 2, pp. 1377–1380 (2001)
3. Yan, Q., Vaseghi, S.: Analysis, Modelling and Synthesis of Formant Spaces of British, Australian and American English Accents. *IEEE transactions on Speech and Audio Processing* 15(2), 676–689 (2007)
4. Yang, D., Iwano, K., Furui, S.: Accent analysis for mandarin large vocabulary continuous speech recognition. *IEICE Tech. Rep.* 107(551), 87–92 (2008)
5. Linqun, L.: Research on A Small Data Set Based Acoustic Modeling for Dialectal Chinese Speech Recognition, PhD thesis, Tsinghua University (2007)
6. Cheng, T.: Huang C., Chang E., Automatic Accent Identification Using Gaussian Mixture Models. In: *IEEE Workshop on ASRU 2001*, pp. 343–346 (2001)
7. Arslan, L.M., Hansen, J.H.L.: A Study of Temporal Features and Frequency Characteristics in American English Foreign Accent. *J. Acoust. Soc. Am.* 102(1), 28–40 (1997)
8. Lin, W., Yang, L., Xu, B.: Speaker Identification in Chinese Whispered Speech Based on Modified-MFCC. *Journal of Nanjing University* 42(1) (January 2006)
9. Li, A., Yin, Z., Wang, T.: RASC863-A Chinese Speech Corpus with Four Regional Accents, ICSLT-O -COCOSDA, India, New Delhi (2004)
10. Wang, W., Deng, H.: Speaker recognition system using MFCC features and vector quantization. *Chinese Journal of Scientific Instrument* 27(6) (June 2006)

# A 24GHz Microstrip Array Antenna with Low Side Lobe

Zhang Zhong-xiang<sup>1</sup>, Chen Chang<sup>2</sup>, Wu Xian-liang<sup>1</sup>, and Fu Ming-gang<sup>3</sup>

<sup>1</sup> Dept. PEE, Hefei Normal University, Hefei, Anhui, 230601, China

<sup>2</sup> Dept. EEIS, University of Science & Technology of China, Hefei, Anhui, 230041, China

<sup>3</sup> No.95028 Air Force, Wuhan, 430079, China

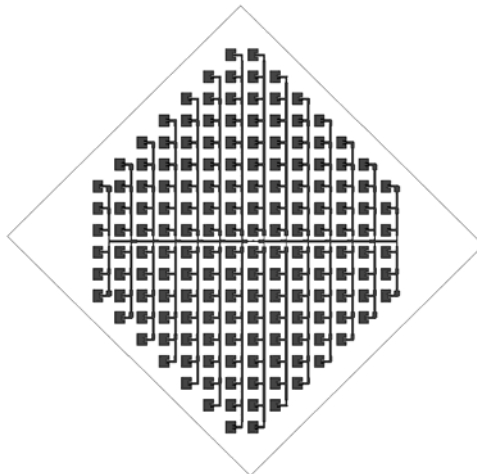
{zhzhx, chench}@mail.ustc.edu.cn, xluwu@ahu.edu.cn

**Abstract.** In this paper, a novel 24GHz 168-element microstrip array consisting of rectangular microstrip resonators is presented. Compared with the traditional 18×14 array, 84 elements may be reduced by keeping same performance. In order to obtain low side lobe, the feeding network with Dolph-Chebyshev distribution is used. And the design is confirmed by simulation and measured results.

**Keywords:** Microstrip array antenna, Low side lobe, Feeding network.

## 1 Introduction

Planar antenna attracts increasing interest due to their flat profile, low weight, case of easy fabrication and low cost [1-4]. In this paper, the design of a simple 24GHz microstrip array is described. This antenna is planned to be integrated in speed radar and intelligent transportation network system.



**Fig. 1.** Layout of microstrip array

The microstrip array shown in Fig.1 is designed on a 10mil thick RT-Duroid 5880 substrate material and consists 168 elements of rectangular microstrip antenna. Compared with the traditional  $18 \times 14$  array, it has the same radiation performance and reduces 84 elements.

## 2 Microstrip Array Simulation

The microstrip array antenna is designed at 24GHz on 10 mil thick RT-Duroid 5880 substrate material, the 3dB beam-widths are  $7^\circ$  in the H-plane and  $9^\circ$  in the E-plane. And the side lobe of the array antenna is less than -17dB.

First step, the single rectangular microstrip resonator shown in Fig.2 is simulated. The size of resonator is  $4.94\text{mm} \times 4.08\text{mm}$  on 10mil- RT-Duroid 5880 at 24GHz.

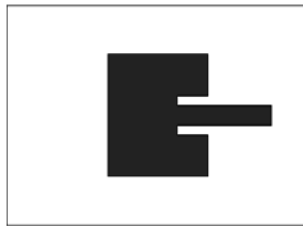


Fig. 2. Structure of microstrip resonator unit

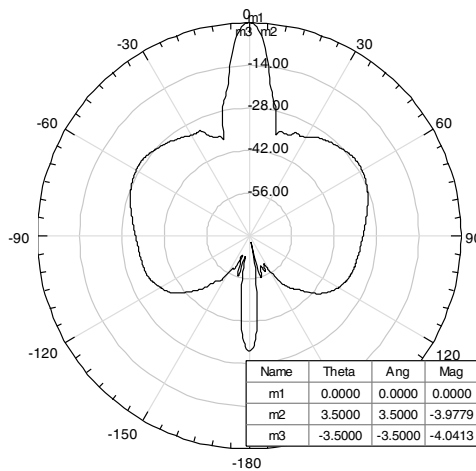


Fig. 3. H-plane radiation pattern of simulation

Second, calculates the feeding network of microstrip array shown in Fig.1. To obtain the 3dB beam-widths in the H-plane and E-plane, the maximum element in the H-plane is 18 and reduces one by one to 6 in the E-plane. Furthermore, the

Dolph-Chebyshev distribution factors and the impedances of the feeding transmission lines are calculated with -17dB side lobe and element numbers in the E-plane and H-plane [5-6].

At last, the whole microstrip array antenna is simulated by EM soft. The H-plane radiation pattern is shown in Fig.3 and 3dB beam-width in the H-plane is less than 7°. The E-plane radiation pattern is shown in Fig.4 and 3dB beam-width in the H-plane is less than 9°.

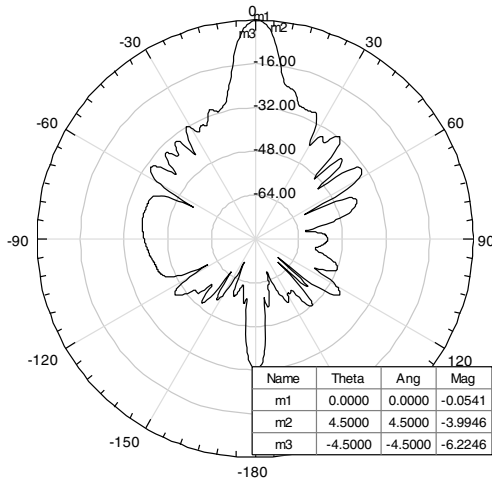


Fig. 4. E-plane radiation pattern of simulation

### 3 Microstrip Array Realization

The microstrip array antenna fabricated on 10mil thick RT-Duroid 5880 substrate material is shown Fig.5.

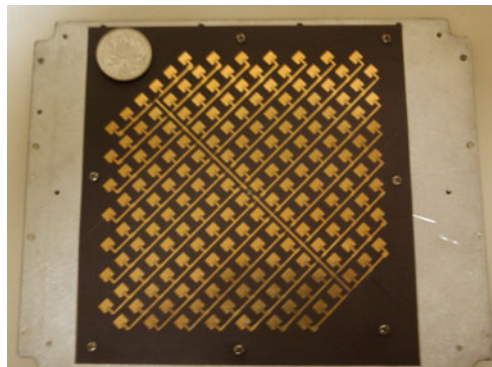
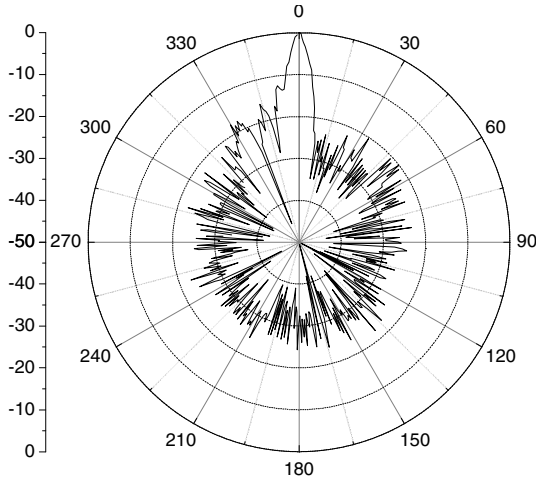
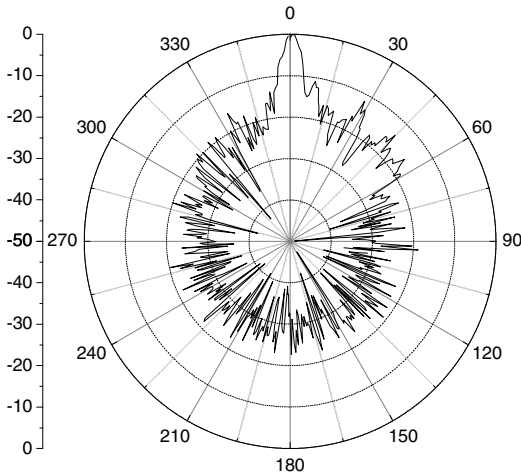


Fig. 5. Photo of microstrip array

Fig.6 and Fig.7 give the measured radiation patterns. The 3dB beam widths of H-plane and E-plane are less than  $7^\circ$  and  $9^\circ$ . The side lobes of H-plane and E-plane are less than  $-14\text{dB}$  and  $-11.5\text{dB}$ . Considering the tolerance of machining, the measured results are consistent with the simulated results.



**Fig. 6.** H-plane radiation pattern of measure



**Fig. 7.** E-plane radiation pattern of measure

**Acknowledgments.** This work is supported by the College Nature Science Foundation of Anhui Province, China (No.KJ2011A241) and 5th Hefei Normal University Foundation (No. 2011rcjj05).

## References

1. James, J.R., Wilson, G.J.: Microstrip antennas and arrays, Pt. 1-Fundamental actions and limitations. *IEE Journal on MOA*, 165–174 (January 1977)
2. Singh, V.K., Singh, M., Kumar, A., Shukla, A.K.: Design of K-band printed array antenna for SATCOM applications. In: *IEEE Proceedings of ICM*, pp. 702–704 (2008)
3. Pozar, D.M., Schaubert, D.H.: Microstrip antennas – the analysis and design of microstrip antennas and arrays. *IEEE Trans. on AP* (1995)
4. Williams, J.: A 36GHz printed planar array. *IEE Electronics Letter*, 136–137 (March 1978)
5. Kim, K.Y., Chung, Y.H., Choe, Y.S.: Low side lobe series-fed planar array at 20GHz. In: *IEEE 1998 AP-S*, pp. 1196–1199 (June 1998)
6. Daniel, J.P., Penard, E., Nedelec, M.: Design of low cost printed antenna arrays. *Institute of Electronics Information and Communication Engineer*, 121–124 (August 1985)



# Resources Allocation of Multi-project in Railway Construction

Zhang Yunning and Jin Yao

Business School of Hohai University, Nanjing, China  
Zhangyunning1959@yahoo.com.cn, Jinyao0919@hotmail.com

**Abstract.** Owing to the amazing rise attention of the government has paid to the railway construction, and high-speed projects have been started in various places, the railway construction enterprises begin to focus more emphasis upon how to implement and balance multi-projects at the same time. On the other hand, these projects occupy the cherish resources and lead to competition. In this research, Genetic Algorithm (GA) is used to solve the allocation resources problem in order to achieve the shortest total duration. Finally, an example is as a proof to the conclusion.

**Keywords:** project scheduling, resources constraints, genetic algorithm (GA).

## 1 Introduction

Nothing can play a more essential role than railway construction in the relationship between national economy and social development, it's the basic condition of production and life of modern society and it's the vital foundation of the modern industry and market-based economic development, what's more, it's the crucial part to achieve the optimal allocation of social resources. After the financial crisis, our country has increased the investment in railway construction, and high-speed rail projects have been started in various places both domestic and foreign countries, so this process tends to face the appropriate allocation of resources, project scheduling and other issues.

Project scheduling as an significant part of the field project management, which means task arrangements during the project under the execution order of each task and the beginning time. The garget of project scheduling is the same as the class of optimization problems. Reasonable project scheduling can shorten the duration of the project and optimize the resource utilization, which in turn to improve the productivity and reduce the total costs [1].

Nevertheless, various kinds of resources are indispensable in the actual project implementation process, such as equipment, workforce, material, but these resources are not infinite supply. The tradition way to solve this problem is CPM or PERT, but these measures often ignore the resources constrains, which leads to resources constrain project scheduling problem (RCPSP)[2]. Typical RCPSP is a study aims to meet the tight project constrains of the former premise, and resource constraints. the task of determining the project starts and end times, and make the shortest project duration. This problem model is to minimize the duration of the project as the goal[3].

## 2 Resources Allocation of Multi-project Problem in Railway Construction

The basic resource-constrained project scheduling problem of the railway construction can be summarized as a general description of the following aspects:

Task:  $A_j$  ( $j=1, 2, \dots, j$ ) denotes a single project contains multiple tasks, and these task must be arranged in a certain order to ensure the smooth implementation of the project. Both  $A_o$  and  $A_j$  are virtual task, which mark the beginning and end of the project, but these tasks do not take up any time and resources. Once the task  $A_j$  begin, it can not be interrupted. Each task has its own duration  $d_j$ ,  $ST_j$  denotes the beginning time and  $ET_j$  denotes the end time. All tasks have two constrains: the tight project constrains of the former premise and resource constraints [4].

The tight project constrains of the former premise: each task must wait until all its predecessor tasks have completed,  $P_j$  as the collection set of the tight project constrains of the former premise. So any task  $h$  belongs to  $P_j$  is not finished, the task  $A_j$  can not be started.

Resource constraints: In order to complete these tasks, resources are consumed, but the capacity of the required resources is limited. The whole task contains many kinds of resources  $k$  ( $k=1, 2, \dots, k$ ), and  $R_k$  denotes the total capacity of the resource  $k$ . In the course of every moment, the amount of task  $j$  consumes resource  $k$  is  $r_{jk}$ , and it can not exceed the total amount of resources.

## 3 Model Assumptions and Notation

The basic optimal goal of RCPSP is to meet the tight project constrains of the former premise, and resource constraints, determine the project starts and end times, and make the shortest project duration. This problem model is to minimize the duration of the project as the goal.

In this model,  $P_j$  and  $r_{jk}$  are assumed as non-negative deterministic parameters. At time  $t$ , the collection of the implementation task is  $I_t$ , so we get following objective function.

$$\text{Min max } \{ FT_{ij} \} \tag{1}$$

$$\text{S.T. } ST_{ij} \geq FT_{ih}, \forall h \in P_{ij}, \forall i, j \tag{2}$$

$$\sum_{A_{ij} \in I_t} r_{ijk} \leq R_k, \forall k, t \tag{3}$$

$$ST_{ij} \geq 0, \forall i, j \tag{4}$$

Especially,  $\max\{FT_{ij}\}$  denotes the end time of the last task of project  $i$ , and it starts at the very begging of process, so it represent the whole duration of the project. As we all know, the objection function is to achieve the minimum. Secondly, the second constraint means the relationship between each tasks of the project. Each task can start after all its procedures have been done. The third constraint means that the amount of task consumes resource can not exceed the total amount of resources at any time during the course. Finally, the forth constraint limits the beginning time of all tasks are non-negative deterministic parameters.

**Table 1.** Basic resource-constrained project scheduling notation

symbol	explain
$i$	Each problem has many independent project $I, i=1, 2, \dots, N$
$j$	Task number, $j=1, 2, \dots, j$ ; set $j$ as the total number of the project
$t$	Time number, $t=0, 1, 2, \dots, t$ ; set $T$ as it up-bound
$k$	Resource number, $k=1, 2, \dots, K$
$d_j$	The duration of task $j$
$P_j$	The collection of the tight project constrains $j$ of the former premise
$S_j$	The collection of the task after task $j$
$ST_j$	The start time of the task $j$
$FT_j$	The end time of the task $j$ and $FT_j=ST_j+d_j$
$I_t$	The collection of the implementation task at time $t$
$R$	The set of renewable resources required for the project
$R_k$	The amount supply of resource $k$ . At any time, resource $k$ can not exceed the total amount $R_k$
$r_{jk}$	The amount of the require resource for task $j$

## 4 Genetic Algorithms (GA)

### 4.1 Generate Initial Population

In order both to ensure the quality of initial population and the efficiency of the algorithm, random way is used to generate initial population. As a precondition to meet the the tight project constrains of the former premise and resource constraints, POPa (the number of the initial generation group) was generated in random way, which means we select from the collection of the tight project set of the former premise, do not use any priority rules. However, POPb (the number of the initial generation individual) choose the next task based on the random distribution priority value mechanism and the minimum time difference between the task priorities. Both the two ways use the serial schedule generation mechanism.

### 4.2 Fitness Function

Fitness value is generated in the decoding of the chromosome scheduling plan which is the sole criterion for evaluation of the merits of the chromosome. Moreover, in the selection operator, we also need to use the size of the individual's fitness to carry out the evolutionary survival of the fittest. In view of the provisions of the genetic algorithm, the fitness value is non-negative value, usually we hope the value of the data as large as possible, but this goal is to minimize the total project duration, part of the minimization problem. So we need to transform the object function to the fitness function in the following way:

$$f(i) = \frac{1}{1+F_i} \tag{5}$$

Among this function,  $f(i)$  denotes the fitness value of the individual and  $F_i$  represent the total duration of the individual  $i$ .

### 4.3 The Design of Genetic Operators

#### Selection Operator

First, before the crossover, parent population must be selected from the current group, and then decide whether to cross or not based on the crossover probability. Therefore, the value of individual fitness is the foundation of the evaluation to select the operator, and then use the select individual to the next step. In the selection method used here is based on the probability of ordering the allocation of options, and then drive under the roulette selection method for individual choice [5].

Secondly, after select the mutation operator, this article will introduce the elite selection strategy, which directly selects POPbest from the previous generation and copy to the next population, without any action on this way to protect a generation of outstanding individuals to ensure the optimal solution is not destroyed. The process is as follows:

Calculate the fitness value of all individuals  $f(i)$ . In the group, the probability select chance of the individual depends entirely on the value of the individual's fitness value.

Determine the selection probability of all individuals as  $P_s(i)$ . Given that using the proportion choice in determining the probability of the select value, the absolute value of the fitness may seriously influence the choice, ie the probability of the more outstanding individuals selected will be far more than poor individuals, resulting in smaller search quickly, and lead to premature convergence phenomenon. Therefore, in this article we used the selection based on ranking selection approach to identify the individual's choice probability, ie the size of the individual to sort according to fitness value and mark the reorder individual as  $\{C_1, \dots, C_n\}$ , so  $f(C_1) \geq f(C_2) \geq \dots \geq f(C_n)$ .

$$P_s(C_i) = \frac{1}{G} \left[ \eta \left( 1 - \frac{\eta^1 - \eta^2}{G - 1} i \right) - 1 \right], \quad i=1, \dots, G \tag{6}$$

Roulette selection method is used for individual choice. And there are needs to calculate the cumulative probability of each individual:

$$PP_s(C_i) = \sum_{j=1}^i P_s(C_j) \tag{7}$$

Each round will generate a number obey uniform random between 0 and 1, this random number, as a pointer, to determine which individual to be elected. In each round, the select-pointer falls on, the cumulative probability distribution of the individual range of the second individual will be selected.

#### Crossover Operator

In this article, point cross be used to calculate, and the probability of it is  $P_e$ . Before each cross will randomly generate a random number  $\epsilon$  and  $\epsilon$  belongs to the range 0 to 1. If  $P_e$  is larger than  $\epsilon$ , the selection operator will select two individual to

cross-operating, and to produce a new individual to conduct mutation operation. Otherwise, it will directly jump into the mutation operation.

CHRM= (αM, βM) denotes the mother individual and CHRF= (αF, βF) denotes the father individual in two parent individuals which will participate the crossover operator. After the crossover operator, offspring are daughter(CHRD=(αD, βD)) and son(CHRS= (αS, βS) ),and the process of cross followed by list α, and daughter sub-project inherited from the mother and the son subprojects inherited the individual priority, that is βD=βM, βS=βF.

Then, a random integer will be generated between 1 and G. The first few positions (r) of the genes of the daughter αD inherited from maternal αM, that is

$$GeiD=GeiM, i=1, 2, \Lambda, r \tag{8}$$

And the rest positions of the daughter (G-r) will inherited from father generation list αF, and the task in the βD will be ignore, and maintain the relative position in the αF, ie.,

$$GeiD=Ge, \tag{9}$$

$$k=\min \{ k|GekF \notin \{ Ge1D,Ge2D,\Lambda,Gei-1D \} , k=1,2,\Lambda,G \} ,i=1,2,\Lambda,r.$$

Similarly, you can get the son individual.

**Mutation Operator**

Each mutation operator operates on only one chromosome; there are two modes of operation: switch and insert [7].

Switching variation which known as exchange the task list between two different tasks, after the exchange, if it does not meet the tight constrains of previous, they will back to the original location and enter into the next exchange [8].

Inserting variation which means: calculate the genetic variation of all predecessor tasks in the task list at the last position (r1) and all after tasks in the task list at the top position (r2) at first. Then, select from r1 and r2 randomly, insert the gene in this position.

As above, the random of the first method, switching variation is too strong, what’s more, it has low success rate, so the mutation operators can not achieve the best consequence, so in this research, inserting variation is used.

**5 Example**

This article assumes a railway construction company contains 5 sub-project assumptions, each sub-project has 15tasks, and all tasks share the 13 kinds of resources. Different projects have different lengths of duration and varying levels of resources. In this paper, a random set of methods is used, and the duration of the project tasks and resource requirements needed are taken as a random integer value between one and nine.

In this paper, this operation involved the use of matlab programming operations to achieve, and by using the GA program, we get the shortest duration is 55.

## 6 Summary

From the perspective of the process of building the railway has many different tasks, this paper expand the basic RCPSP model, and focus on the railway construction company face multiple schedule problem under resource constraints. It's a good solution to use GA to solve the multi-project resource scheduling problem. However, in practice, many parameters setting is uncertain, the future can be improved from the perspective of the simulation algorithm. Secondly, the purpose of this study is to make the total number of projects the shortest duration, but may lead to a single project to extend the duration, so there is much to improve in this place.

**Acknowledgments.** AS a graduate student, I appreciate my supervisor, he support me to finish this paper. And also he gives me money to publish it. Meanwhile I want to thank the conference which gives me a chance to publish my paper.

## References

1. Liu, Y.-M., Wang, Z.-F., Zhang, Y.-M.: Time-resources tradeoffs of PERT construction schedule under resource constraints. *Advances in Science and Technology of Water Resources*, 27–30 (2007)
2. Duan, G.-L., Zha, J.-Z., Xu, A.-P.: The applications of heuristic algorithms in engineering. *Machine Design*, 1–5 (2008)
3. Gan, T.-A., Tian, F., Li, W.-Z.: *Operational Research*, 3rd edn. Pek University, Beijing (2008)
4. Zhao, C.-Y., Zhang, L.: A Good-point Set Genetic Algorithm for TSP. *Computer Engineering and Applications*, 103–105 (2008)
5. He, J.-Y., Bai, S.-Q.: Research on critical chain schedule optimization. *Management and Practice*, 33–36 (2008)
6. Zhang, Y.-N., Sun, X.-X.: Research on Improved PERT Model Application in Analysis of Schedule Risk of Project. *Science and Technology Progress and Policy* 25(10), 94–96 (2008)
7. Zheng, X.-L., Li, F., Di, P.: Research on Application of Integral Programming PSO Algorithm to Spare Part Optimizing. *Ship Electronic Engineering* 29(3), 133–135 (2009)
8. Zhang, H.-P., Qiu, W.-H.: A Modified Genetic Algorithm for Scheduling Multiple Projects with Resource Constraint. *Chinese Journal of Management Science* 15(5), 78–82 (2007)

# The Research on Development Risk of Green Construction Material Based on Fuzzy Comprehensive Evaluation

Zhang Yunning and Liu Gang

Business School of Hohai University, Nanjing, China  
Zhangyunning1959@yahoo.com.cn, ganggangliu@126.com

**Abstract.** In order to analyze and select the development project of green construction material more scientific and rational, the fuzzy comprehensive evaluation method in fuzzy set theory is introduced in this paper to evaluate development risk and establish comprehensive evaluation model for development risk of green construction material. This paper adopts risk priority number (RPN) which from failure mode effects and criticality analysis (FEMCA), to analyze the risk scale, set risk weight, and let this setting closer to actual condition, then calculating examples. At last we conclude that the method provides feasible analysis foundation, and has the strong practical value.

**Keywords:** green construction material, risk, fuzzy comprehensive evaluation, femac.

## 1 Introduction

In recent years, the public voice for environment protection is increasingly rising, people advocates green building and sustainable development, as the vector of green building, the development of green construction material has become the focus of building materials industry[1].The study of green construction material began in the 1990s and have made certain achievements[2].But there are various problems on green construction material which by independent development in our country, The effect of the popularization and application of the product also non-ideal. How to avoid these problems, selecting and developing a building materials scientifically, and applying to architectural entity successfully, gain profit, It's need to think over in green construction material filed. To solve the problem, it's an effective method to make risk analysis on the developing of green construction material. Through researching the developing risk, we can select the development projects reasonably and improve the success rate of development.

At present, the research on development risk of green construction material is less. They are involved the following several aspects: 1.Market risk and market development, 2. Unsoundness of policy and regulation, 3. Resource materials problems; 4.Operation and management problems; 5.Technical risks; 6.Other problems, such as management model problems etc. While basing on the above

research and related theories on fuzzy mathematics, this paper constructed fuzzy evaluation model of development risk on green construction material, studied the evaluation for development risk of green construction material.

## 2 Fuzzy Evaluation Model for Development Risk of Green Construction Material

### 2.1 Establishing Risk Factors Structure

Through reading a lot of literature, collecting and arranging data, based on the different attributes of development risk of green construction material, we can divide the evaluation objects of development risk of green construction material into three levels, technical risk level, economic risk level, social environment risk level. Let risk assessment factor set of total green construction material risks be a first risk factor set  $R$ , then  $R$  contains three first risk factors and  $R=(R_1,R_2,R_3)$ , meanwhile, every first risk factor contains several second risk factors, the concrete conditions in the table 1:

**Table 1.** Green building materials development risk factors structure

Indexes of first risk factors $R_i$	Indexes of second risk factors $R_{ij}$
<b>Technical risk <math>R_1</math></b>	Design risk $R_{11}$
	Innovation risk $R_{12}$
	Materials risk $R_{13}$
	Effect risk $R_{14}$
	Resource risk $R_{15}$
<b>Economic risk <math>R_2</math></b>	Investment risk $R_{21}$
	Market risk $R_{22}$
	Price risk $R_{23}$
	Operation mode risk $R_{24}$
	Income risk $R_{25}$
<b>Social environment risk <math>R_3</math></b>	Policy risk $R_{31}$
	Environmental protection Consciousness risk $R_{32}$
	Understanding of environmental protection $R_{33}$
	Cooperation risk $R_{34}$
	Evaluation risk $R_{35}$

### 2.2 Determining Comment Set

According to the above risk system, the risk level can be divided into five grades: higher risk, high risk, medium risk, low risk, lower risk, they are corresponding these digital respectively: 1,2,3,4,5. Then we get the comment set  $V=\{1,2,3,4,5\}=\{ \text{higher risk, high risk, medium risk, low risk, lower risk} \}$ .

### 2.3 Establishing Corresponding Weight Set

According to the above corresponding evaluation factor set, we established weight set respectively: First weight set  $A$ ; Second weight set  $A_1, A_2, A_3, A_4, A_5$ ; Among them, Tab.1 cutting



$$A=[w_1, w_2, w_3](i=1,2,3), A_i=A_i[w_{i1}, \dots, w_{ij}], (i=1,2,3; j=1,2,3,4,5) \text{ and } \sum_{i=1}^3 W_i = 1$$

$\sum_{j=1}^5 W_{ij} = 1, W_i \geq 0, W_{ij} \geq 0$ . The key of fuzzy comprehensive evaluation is how to

distributing the weight coefficient. the occurring probability and consequences caused are difference due to different risk factors, to make evaluating results more scientific and real. This paper adopts risk priority number (RPN) which from failure mode effects and criticality analysis(FMECA), to analyze the risk scale and set the weight of every factor. Among them,  $RPN = \text{score of occurrence degree} \times \text{score of severity} \times \text{score of detection degree}$ . The score of occurrence degree, severity and detection degree are all in the interval (0 , corresponding weight. Because the traditional RPN method require the score of occurrence degree, severity and detection degree must be very accurate, or it'll lead to larger floating for the result, so this paper introducing fuzzy logic RPN, using blurring, 10), so  $0 \leq RPN \leq 1000$ , the larger of the score of RPN ,the larger of the fuzzy reasoning to treat the input occurrence degree  $I_{OX}(O)$ , severity  $I_{SY}(N)$ , detection\_degree  $I_{DZ}(D)$ , pushing out the value of risk degree  $IRW(R)$  to overcome this defect. Suppose the universe of  $I_{OX}(O), I_{SY}(N), I_{DZ}(D)$  are {lower, low, middle , high, higher}, and the universe of the value of risk degree  $IRW(R)$  are {smaller, small, middle, big, bigger}. After treatment, we can obtain the corresponding fuzzy interval respectively, they are: Fuzzy-valued of occurrence degree: lower[0,2], low[1,4], middle[3,7], high[6,9], higher[8,10]; Fuzzy-valued of severity: lower[0,2], low[1,4], middle[3,7], high[6,9], higher[8,10]; Fuzzy-valued of detection degree: higher[0,3], high[2,5], middle[4,7], low[6,9], lower[8,10]. Fuzzy-valued of risk degree: smaller[0,200], small[0,500], middle[200,700], big[500,1000], bigger[800,1000].

According to the advices of experts, we obtain corresponding fuzzy inference rules. If occurrence degree is a little low, severity is high, and detection degree is middle, then the risk degree is a little big. Arrange these rules to be corresponding rule base. The comprehensive rule base of fuzzy inference from fuzzy logic RPN rule is as the table 2 shows:

**Table 2.** Comprehensive rule base of fuzzy risk degree

rules	occurrence degree	severity	detection degree	risk degree
1	low	middle	lower, low, middle	middle
2	low	higher	lower, low, middle	big
3	low	higher	lower, low	bigger
4	low	higher	middle	big
5	middle	middle	lower, low	middle
6	middle	higher	lower, low	bigger
7	middle	higher	middle, high, higher	big
8	middle	very high	lower, low, middle	bigger
.....				

We adopt operator  $M=(\bullet, +)$ , using knowledge of fuzzy mathematics to obtain the value of risk degree  $IRW(R)$  respectively which corresponding with risk factors of every level, then use weighted normalization to processing to weight of every risk factors.

**2.4 Establishing Evaluation Model of First Risk Factor**

According to the high-low of risk degree of every risk factors, the experts of this filed evaluate membership vector  $r_{ij}$  of every second risk factor  $R_{ij}$  the risk evaluation set  $V$  by reference to the evaluation standard. The membership matrix is:

$$R_i = \begin{bmatrix} r_{i1} \\ r_{i2} \\ \dots \\ r_{ij} \end{bmatrix} = \begin{bmatrix} r_{i11} & r_{i12} & \dots & r_{i15} \\ r_{i21} & r_{i22} & \dots & r_{i25} \\ \dots & \dots & \dots & \dots \\ r_{ij1} & r_{ij2} & \dots & r_{ij5} \end{bmatrix}$$

$r_{ijv} = u_{ijv} / N$ ,  $N$  is the Total number of experts,  $u_{ijv}$  is the number of people who consider  $R_{ij}$  is belong risk grade  $v \{v = (1, 2, 3, 4, 5)\}$ . We can obtain the fuzzy comprehensive evaluation model of  $R_i$  is  $B_i$

$$B_i = W_i \bullet R_i = (W_{i1}, W_{i2}, \dots, W_{i5}) \bullet R_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4}, b_{i5}) \tag{1}$$

Among them,  $B_i$  is the membership matrix for first risk factor  $R_i$  to risk evaluation

set  $V$ ,  $b_{ij} = \sum_{j=1}^5 W_{ij} u_{ijv}, v = (1, 2, 3, 4, 5)$ ,  $W_{ij}$  is weight vector of every level's second risk

factor.

**2.5 Establishing Comprehensive Evaluation Model of Risk System**

According to the obtained  $B_i (i = 1, 2, 3)$ , we can get the membership matrix of single

factor (subset) of  $R$ : 
$$R = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} =$$

Till here, we obtain the fuzzy comprehensive evaluation model of  $R$ , namely risk comprehensive evaluation model of one kind of green construction material:

$$B = W \bullet R = [W_1, W_2, W_3] \bullet \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \end{bmatrix} = [b_1, b_2, b_3, b_4, b_5] \tag{2}$$

Based on the above model, the Evaluation conclusion of R can be obtained by maximum membership degree method. If the result  $b_m = \max(b_1, b_2, b_3, b_4, b_5)$ , then we can evaluate R is m level. But if there are two or more than two component equal, then this method lose efficacy.

### 3 Example of Model Application

#### 3.1 Model Instance Calculating

There is a close relationship between the using of building material and the indoor air quality, especially on today when chemical industry tied up with building material. The writer based on one development organization's project (This project try to develop a decoration material which can enhance the indoor air quality), using models established to assess the project development risk. The specific process of evaluating the develop risk is as below:

First, establishing the risk factors system of green construction material R(show as the table 2-1) and evaluation set  $V = \{1,2,3,4,5\} = \{\text{risk is very high, risk is high, risk is middle, risk is low, risk is very low}\}$ . Secondly, consulting the relevant expert, and according to the concrete conditions occurrence degree, severity, detection degree, using RPN which in fuzzy logic method, we obtained the weight of every level's factors, shown as below:

$$W = [w_1, w_2, w_3] = [0.4, 0.3, 0.3]$$

$$W_1 = [w_{11}, w_{12}, w_{13}, w_{14}, w_{15}] = [0.3, 0.15, 0.15, 0.15, 0.2]$$

$$W_2 = [w_{21}, w_{22}, w_{23}, w_{24}, w_{25}] = [0.3, 0.3, 0.1, 0.2, 0.1]$$

$$W_3 = [w_{31}, w_{32}, w_{33}, w_{34}, w_{35}] = [0.3, 0.15, 0.15, 0.25, 0.15]$$

Subsequently, we adopt the suggestion of every risk factor which made by experts, and obtained the membership matrix of every second risk factor to first risk factor, showed by  $R_1, R_2, R_3$  respectively, which as below:

$$R_1 = \begin{bmatrix} 0.1 & 0.4 & 0.3 & 0.2 & 0 \\ 0 & 0.4 & 0.3 & 0.3 & 0 \\ 0 & 0.1 & 0.3 & 0.5 & 0.1 \\ 0 & 0.1 & 0.2 & 0.5 & 0.2 \\ 0 & 0.2 & 0.2 & 0.5 & 0.1 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0 & 0.3 & 0.3 & 0.4 & 0 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.2 & 0.3 & 0.5 & 0 \\ 0 & 0.1 & 0.2 & 0.6 & 0.1 \\ 0 & 0.1 & 0.1 & 0.6 & 0.2 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.2 & 0.3 & 0.4 & 0.1 \\ 0 & 0.3 & 0.4 & 0.3 & 0 \\ 0 & 0.1 & 0.2 & 0.6 & 0.1 \end{bmatrix}$$

Using formula (1), we can obtain fuzzy comprehensive evaluation of  $R_1, R_2, R_3$  :  
 $B_1 = [0.03, 0.26, 0.26, 0.38, 0.07]$ ;  $B_2 = [0, 0.17, 0.2, 0.56, 0.07]$ ;  $B_3 = [0, 0.135, 0.19, 0.33, 0.345]$

Using above results, we obtain the membership matrix of first risk factor in R, then through formula (2), we obtain the risk fuzzy comprehensive evaluation:

$$B = W \cdot R = [w_1, w_2, w_3] \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \end{bmatrix}$$

$$= [0.012, 0.196, 0.221, 0.419, 0.153]$$

Though the front calculating, we know the evaluation results of risk fuzzy synthetic in R, seen as below:  $B = (0.012, 0.196, 0.221, 0.419, 0.153)$ . According to maximum membership body of the paper here, degree method: because  $B = (0.012, 0.196, 0.221, 0.419, 0.153)$ , estimating through maximum membership degree, the result of this development project’s risk assessment is “Risk is a little low”.

Besides, according to the suggestion of expert, we must pay sufficient attention to the factors which bigger than 0 in  $r_{ij}$ . It explains that this factor is unacceptable in this development project. In this project, the risk which corresponding  $r_{ij}$  and bigger than 0 is innovation risk. So in the project development, measures must be adopt and reasonable strategy must be making to control the risk.

## 4 Summary

This paper adopts comprehensive fuzzy evaluation method to establish the risk evaluation model of green construction material development, and make analysis on corresponding practical example. We can use this model to make quantitative evaluation and analysis on the risk in the early stage of the project, and providing an effective scientific decision and risk management tool for development organizations. Meanwhile, to develop a kind of green construction material successfully, we should pay sufficient attention to market and technology.

**Acknowledgments.** AS a graduate student, I appreciate my supervisor, he supports me to finish this paper. And also he gives me money to publish it. Meanwhile I want to thank the conference which gives me a chance to publish my paper.

## References

1. Hu, J.: Exploration and practice of constructing the modern green building system. Chongqing University (2006)
2. Bai, S.P.: Overview of the Green Building Materials. Tech Information Development and Economy (35) (2007)
3. Zhu, H.H.: Resear of green building materials market development in China. Wuhan University of Technology (2002)
4. Wang, H.L., Ding, Y.G., Zhen, J.S.: The philosophical thought of new material and social economic and technical development of the relationship. Science Technology and Dialectics (02) (2001)
5. Zhu, L., Liu, Z.Y.: The analysis on Enterprise green management model. Journal of Hefei University of Technology (Social science edition) (01) (1999)
6. Wang, X.Y., Wang, P.M.: The research and application of green building materials. China building materials industry press (2004)
7. Zhang, J.J.: Fuzzy Analytical Hierarchy Process(FAHP). Fuzzy Systems and Mathematics 6 (2000)

# Design and Development of Provincial Master Station of Metering Automation System of Guangdong Power Grid

Yang Jinfeng, Chen Ruimin, Xiao Yong, Dang Sanlei, and Wei Fusheng

Electric Power Research Institute of Guangdong Power Grid Corporation  
510080, Guangzhou, China  
csjfyang@gmail.com

**Abstract.** In recent years, Guangdong Power Grid Corporation has built the largest scale electric energy metering automation system in the country. The province's 21 Power Supply Bureau have been building master metering automation systems. Because the data scattered in the 21 Power Supply Bureau's master station, the data can not be the province's electricity supply integrated display. There is an urgent need to construct a provincial master station, get the Power Supply Bureau's master station data, process and integrated display the province-level data. In this paper, considering the actual situation of Guangdong Power Grid Corporation, we design and develop a Provincial master station of metering automation system of Guangdong Power Grid. We use advanced SOA and J2EE architecture, and apply a unified data format, get the power supply bureau's master station data, calculate totals and overall show. After two years development and construction, Provincial Master Station of Metering Automation System of Guangdong Power Grid achieved good effect for the Guangdong Power Grid production and marketing business.

**Keywords:** Metering Automation System, Provincial Master Station, J2EE, SOA.

## 1 Introduction

In recent years, with the rapid economic and social development, the serious energy shortages and power supply structural contradictions have become increasingly prominent, and demands for power supply reliability have been increasing, and the service demand of power users have been more diverse. All these urgently requires power grid to a more modern, intelligent direction. Advanced Metering Infrastructure (AMI) is a solid foundation to achieve "informatization, digitization, automatization, interactivation" of power supply and consumption part of smart grid. Comply with the smart grid and AMI development trends, Guangdong Power Grid Corporation (GDPGC) firstly constructed metering automation system in large-scale, installing a large number of metering automation terminals and intelligent electricity meters at power generation, transmission, distribution, and consumption part, achieving

electricity information collection of all part of metering point. Metering automation system has functions of remote meter reading, power load control, online monitoring of metering device, statistical analysis of line loss, statistical analysis of power supply and distribution, providing advanced technical support for orderly power consumption and demand side management.

Guangdong Power Grid's metering automation system has been constructed respectively by various power supply bureau under the guidance of a unified technical specification. There are 21 sets of municipal master stations of metering automation system, constructed by 5 different software development corporations respectively. From the perspective of the province, the data scattered in the master station of 21 Power Supply Bureau, and can not be the province's electricity supply integrated display. Therefore, there is an urgent need to build a provincial master station, gaining data of 21 sets of municipal master stations for the province-level data processing and integrated display.

## 2 Construction of Metering Automation System of GDPGC

In order to ease the contradiction between power supply and demand, strengthen demand side management, and improve service level, GDPGC promoted to construct metering automation system from 2005. After several years of construction, the system has been in large-scale. All 21 power supply bureaus have built a municipal master station, and have accessed more than 400,000 metering automation terminals and 3.5 million metering points. Metering automation system collects data from power plants, substations, special transformer, common transformer, and low voltage users, and has functions of remote meter reading, power load control, online monitoring of metering device, statistical analysis of line loss, statistical analysis of power supply and distribution, providing advanced technical support for orderly power consumption and demand side management.

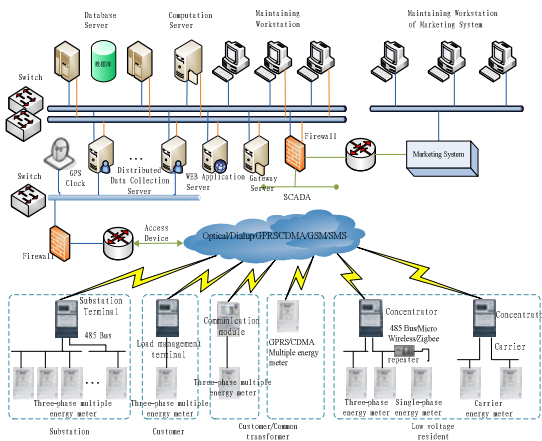
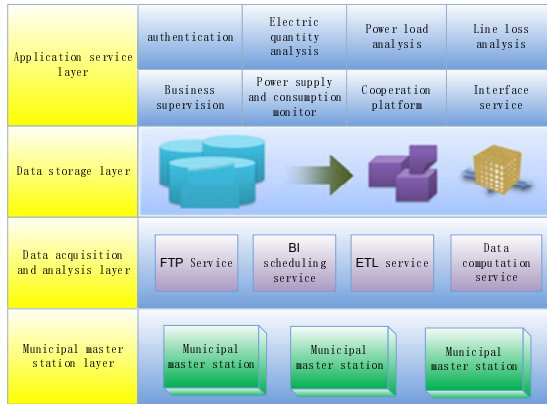


Fig. 1. Architecture of metering automation system







**Fig. 3.** Logical structure of provincial master station

1) Municipal master station layer

21 sets of municipal master stations are data source of provincial master station, and the data includes detailed electric quantity and load of all metering points and summarized power supply and consumption of the power supply bureau.

2) Data acquisition and analysis layer

Municipal master station extracts data to interface server according to relevant interface standard, and provincial master station gains data from interface server by FTP service, and then resolves the data to store in the database.

3) Data storage layer

This layer store the origin data uploaded by municipal master station and analytic data computed by provincial master station.

4) application service layer

This layer provides functions like power load analysis, electric quantity analysis, line loss analysis for business department, and monitors power supply and consumption of the province in real time, providing decision support for power grid production and marketing.

**3.2 Interface Design**

Provincial master station gains the data from 21 sets of municipal master station through interface, since there are differences in database design of the municipal master station developed by different software company, the data format must be unified.

1) Data interchange format

Provincial and municipal master station uses XML files as a data interchange carrier. The delivering content is organized into XML format, such as the district data format:

```
<AMSB rows = "1">
  <townId></townId>
  <townName></townName>
  <powerBureauId></powerBureauId>
  .....
```

```

</AMSB>
.....
<AMSB rows = "n">
  <townId></townId>
  <townName></townName>
  <powerBureauId></powerBureauId>
.....
</AMSB>
    
```

2) Data interchange strategy

Provincial and municipal master station interchanges data using FTP transmission. Provincial master station is a server, and the municipal master station is a client. FTP commands conform to RFC standards, and support breakpoint resume.

3.3 Technical Architecture

1) J2EE Enterprise Application Framework

Provincial master station uses J2EE enterprise application framework, as shown in figure 4.

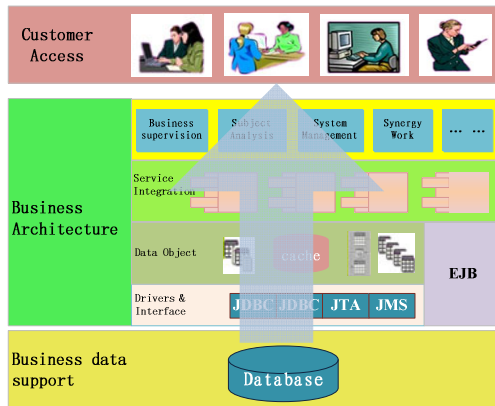


Fig. 4. Technical Architecture of Provincial Master Station

The services provide its users with transparent access to applications. J2EE divides enterprise application into three layers: customer access layer, business architecture layer, and business data support layer.

Customer access layer mainly integrates HTML script application and rich client component of the browser. In this layer, common techniques include HTML, CSS, Javascript, Ajax .etc.

Business architecture layer is the core layer of J2EE. It's the biggest piece of change in building large complex application. Enterprise-class applications are generally built on the J2EE middleware application framework to provide services.

Business data support layer provides data. The data is not only present in the metadata of relational database, but also in other existing applications deployed on SOA bus.

2) Service Oriented Architecture

As there is large number of metering automation terminals and the data collecting frequency is high, the data collected by all municipal master stations is over 100 gigabytes. The provincial master station is faced with problems of massive data storage and processing.

In the process of dealing with massive data, due to its massive and decentralized features, we must establish a distributed application framework to support. In this distributed framework, there is a high degree of business integration in business processing among different business units. In order to ensure the consistency of common information, it's necessary to design independent model of different kinds of basic businesses. In this multi-services and distributed system, there is a need to use idea of SOA architecture to establish a scalable module with low redundancy, high consistency, high availability, and interoperability. Figure 5 shows the technical framework model using the idea of SOA based on Guangdong Power Grid.

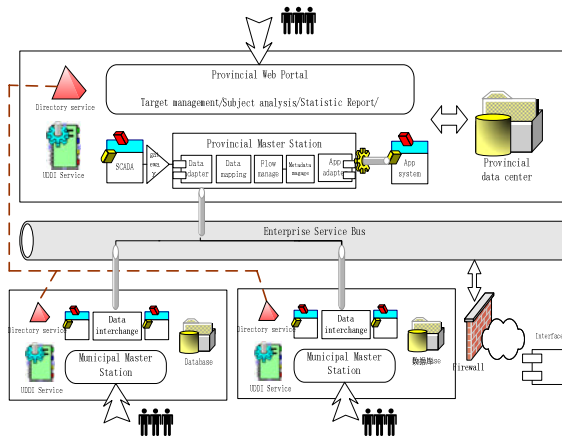


Fig. 5. The technical framework model

3.4 Functions Design

According to business needs, provincial master station mainly includes functions of operation maintenance, subject analysis, business supervision, cooperation work.

Operation maintenance mainly includes functions as data summarizing computation, interface monitoring, and data backup management .etc. It provides technical means to ensure the stability of master station.

Subject analysis provides the power supply and consumption, power load, and line loss analysis and comprehensive multi-dimensional drill-down display.

Business supervision monitors the maintenance status by monitoring some key rate like terminal cover rate, terminal online rate, and so on.

Cooperation work provides processing flow of abnormal data. For abnormal data, maintenance personnel can generate work orders, and inform relevant personnel by SMS, completing closed-loop processing in provincial master station.

## 4 Development and Application

After 2 years construction, the provincial master station has access up to 2239 substations, 266916 special transformers, 189263 public transformers, and 3127207 low voltage domestic users. Marketing department of GDPGC uses the provincial master station to compose daily and monthly report of metering automation system, and supervise the operation and application effect of municipal metering automation systems, and has promoted effectively the main indicators of metering automation system. Besides, the data provided by provincial master station has been applied at production and technology department, and planning department, providing advanced technical support for power grid operation and marketing.

## 5 Conclusion

With the rapid development of society and economic and electricity marketization, demand side management well gain more and more attention from all sectors of society. As one of important basic systems of demand side management, metering automation system gains rapid construction nationwide. It plays an important role at automatic meter reading (AMR), electricity utilization monitoring, power load forecasting and analysis, and line loss analysis and management. The construction of provincial master station resolves the problem of “island of information” of different municipal master stations, and effectively integrates the data from municipal master stations, providing technical and data support for power grid production and marketing.

## References

1. Min, T., Yan, X., Li, J., Li, J., Wang, J.: A New Generation of Electric Energy Metering System Based on J2EE Architecture. *Automation of Electric Power Systems* 27, 85–88 (2003)
2. Zhong, S.: Design and Application of Metering Automation System of Huizhou Power Supply Bureau. *China High Technology Enterprises* 15, 67–69 (2009)
3. Yu, Z., Wu, Y.: Description on Design of Electric energy metering system of Jiangsu Power Grid. *Automation of Electric Power Systems* 30, 106–107 (2006)
4. Huo, X., Li, J., Wu, Y., Yan, X.: The Unified Interconnection Scheme of Energy Metering System Based on E Language. *Automation of Electric Power Systems* 30, 73–76 (2006)

# The Research of Acceptance and Scheduling Scheme Based on the Overall Network QoS

Xu Zhanyang<sup>1,2</sup> and Huang Li Ping<sup>1</sup>

<sup>1</sup> Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, China

<sup>2</sup> School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, China

**Abstract.** Many factors affect the QoS of network, and the congestion control mechanism is the most important one. In many references, the research on QoS is focused on the specific network architecture, while the QoS of overall network is still relatively seldom studied. In the paper, an acceptance and scheduling scheme based on the overall network QoS (ASSoQoS) proposed is a new acceptance and scheduling algorithm of the link traffic. At first, the data stream with the admission control strategy will be distributed in different weights, the data flow will get the link (with different weight) transmission scheduling to improve the overall network QoS, according to the allocated weight of a network link and status by real-time network monitoring.

**Keywords:** Acceptance, scheduling, QoS.

## 1 Introduction

With the development of computer and network communication technology, IP-based network integration and infrastructure have become the inevitable trend of network evolution. The emergence of multi-business contained in network convergence has been required the increase of a higher QoS [1]. Although the traditional model of integrated services can provide end to end Integrated Services (IntServ) for absolute QoS guarantee, but its drawback is increasingly prominent, mainly such as, (1) the resource reservation is based on each stream, the node needs to retain per-flow state information, resulting in overloading of the core router, so scalability is poor; (2) each node must maintains various databases and complex functional blocks (such as resource reservation, routing, admission control etc), which makes the model very complex coefficients. Against with the drawbacks of IntServ model, DiffServ model is proposed by IETF. It is focused on grouping the IP with different quality of service requirements by border router, then different IP group will be assigned a different tag and similar requirements of business will be attributed to a category [2].

The acceptance and scheduling scheme based on the overall network QoS is proposed, and the data flow based on the scheme will improve the overall network QoS.

## 2 Acceptance and Scheduling Algorithm of a New Type Link Traffic

As we all know, network topology is very complex in the real life, abstracted as a way similar to the logistics operation. As a reference, a simple logistics and transportation link model is shown in Fig.1 (seen in Appendix A ). The originating data stream A is divided into different type A1、 A2, which will be sent on different transmission links as in Fig. 1 (which links have different rate for different data flow). How to choose an optimal path for the data flow, completing the transfer task on demand and improving overall network QoS, is focused in the paper. In such case, for a data stream selecting the path, the scheduling algorithm will schedule data transmission stream according to the weight value, achieve the maximum possible link rational use of resources, to improve overall network QoS.

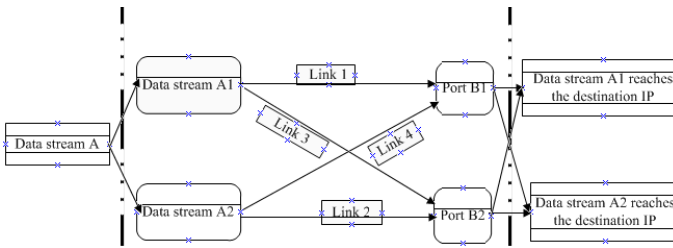


Fig. 1. An reference to the simple logistics and transportation link model

As described above, mainly, there are two kinds of QoS service model, both have strengths and weaknesses, the idea of which are embraced into this scheme. A new transport link to accept scheduling algorithm is proposed which is divided into three parts, namely, data acceptance, data logistics transmission, queue scheduling algorithm and guardian.

### 2.1 Data Acceptance

As references of QoS issues in different networks, the admission control strategy is an important part of providing QoS [3]. Firstly, based on a consideration for admission control strategy, the idea of weight distribution has been introduced in this scheme. This distribution is based on different data streams which are divided into different weight. Specific steps are as follows:

(A) Call a network line data trigger, and pull up the data weight control side. When there are data to transmit in this link, a data trigger is loaded and the weight control side will be triggered. Since polling needs to consume network resources, therefore, it's not recommended to use any way. Then, when this link has no data to transfer, the flip-flop is in a dormant state to maximize savings network resources;

(B) Distribute weights dynamically for data stream accepted and the acceptance controller depending on the regional distribution of different areas. Regional controllers are controlled by the general allocation and monitoring management. For example, when a data transmission link is requested for a data stream, initial data

reception is given as acceptance a1, and the weight of it will be noticed to acceptance a2. If the data transmission link is relatively complex, the assignment will be based on the data destination side of the sub-section, then the weight of acceptance a1 is noticed step by step;

(C) Give different partition acceptance controller in different link-side with different weight distribution to ensure that the link weight distribution is fair according to the total acceptance controller. Fig. 2 shows the data accepted flow diagram.

**2.2 Data Logistics Transmission**

As described in above, firstly, a reasonable weight value distribution is given to the overall network, including data stream weights, control weights, network link weights, regional network weights and so on. Through the whole network monitoring and local network monitoring, the overall network topology link structure can be calculated [4]; then, different exchange routes are able to calculate, prioritize and a performance arrangement is given to them; link weights are calculated through an integrated algorithm which is to guide data streams transmission by a path guidance model. For different data streams, the admission control strategy described above has been taken in for them which are sent from client to divert. Corresponding weights assigned to each stream which is based on the influenza known to distinguish [5] are calculated according to the weight calculation method. As distinguished service to mark different streams, flows with different weights in the scheduling will let scheduling record its own weight (scheduling algorithm is calculated based on the data stream weight and network link line weight); then a link for the data flow is shown; meanwhile the data flow is guided by the intelligent device gps, and controllers jump to next according to its weight. Finally the weight checks at the regional emissions for the queue discharge. If a real-time data transmission is in the progress, under the condition of link carrying capacity, to ensure transmission quality, the link segmentation approach is adopted for passing to take advantage of network resources, according to current network conditions and the most optimal link allocation. The data accepted flow diagram is shown in Fig.2.

(A) Calculate Weights: Weight calculations include data flow weights, control weights, network link weights, and network area weights. It is assumed, the corresponding factors are  $\lambda_s$ ,  $\lambda_c$ ,  $\lambda_l$ ,  $\lambda_a$ , and the other weighting factor is  $f(\lambda)$ , at  $t_i$  second, each weight factor is calculated as flowing :

The data stream weighting factor: A data flow reached to the admission control port at  $t_i$  second, is data flow  $D(i)$ , while the size of this data stream is  $N(i)$ , then, weighting factor of this data flow is (1):

$$\lambda_{si} = N(i) \frac{D(i)}{\sum_{i=1}^{i=n} N(i)D(i)} \tag{1}$$

Controller weighting factor: Suppose that n controllers have already been in the network, when a data request message is sent, the user's source port and destination port have been known through data collection devices, then a network link formed.

Through the router or switch self-forwarding, of course, we can get m different kinds of links. Each of them, we get the number of the existence controllers, assuming that the total number of controllers in i link is N, then  $\lambda_{ci} = 1 / N_i$ .

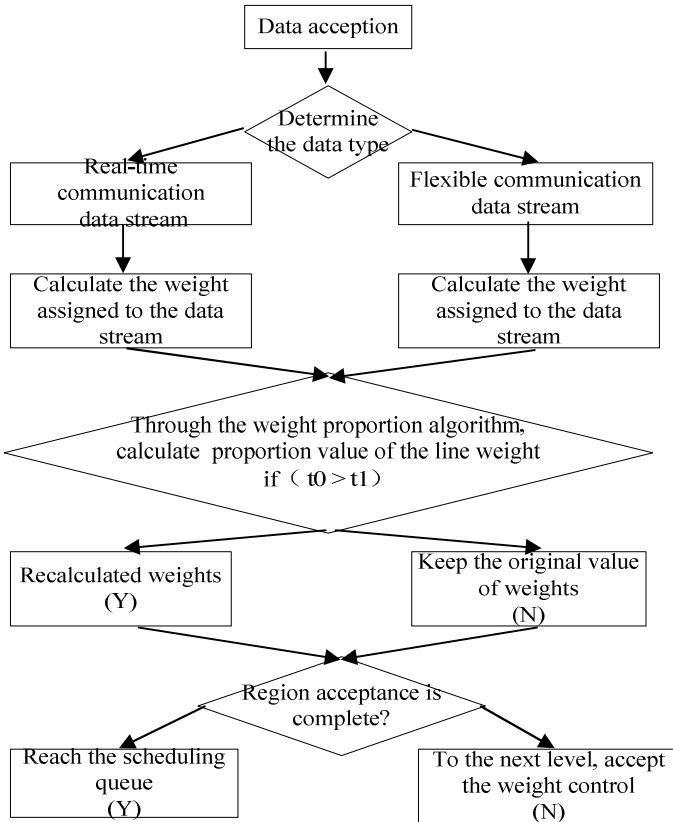


Fig. 2. The data accepted flow diagram

Regional network weighting factor: Dynamic level-hierarchical thinking has been used in this factor. Firstly, assuming that a link between the source IP and destination IP is known, the area which the link maybe go through is also known. Then through the topology discovery, the core of the device region can be discovered (the core of the device is not necessarily one, mainly divided according to type). Secondly, the core network equipments are regarded as the first level, which has the highest weight  $\lambda_{a1}$  meanwhile, the sub-regional level compared with the weight  $\lambda_{a2}$ , and so on. In the i-level, weighting factor of the area is  $\lambda_{ai}$ :

$$\lambda_{ai} = \frac{N(i)\lambda_i}{\sum_{i=1}^{i=n} N(i)\lambda_i} \tag{2}$$



There,  $N(i)$  is the number of devices for the  $i$ -level area,  $\lambda_i$  is for the corresponding weight of each device, which is a good pre-requirements, mainly based on the performance model to distinguish. In general, the higher performance or the higher price of equipment at this level, the greater the weight.

Network link weight factor: Whether the data stream is accepted is not only depends on the calculation above, but also needs to consider the entire network link weight factor. Assuming that, the different link bandwidth is  $B_i$ , link bandwidth level is  $G_i$  (the higher of link bandwidth, the greater the level) [6], then weighting factor for the link in the  $i$  segment is  $\lambda_{li} = l / (B_i G_i)$ , there,  $l$  is the proportional coefficient, which can be achieved through the data flow collection under long-term calculation.

In conclusion, the weight value of data flow  $i$  is  $P(t,i)$  as follows in (3):

$$P(t,i) = \alpha(t)\lambda_{si} + \beta(t)\lambda_{ci} + \chi(t)\lambda_{ai} + \delta(t)\lambda_i + f(\lambda) \tag{3}$$

$\alpha(t), \beta(t), \chi(t), \delta(t)$  are the time scale factors, the other data stream  $j$  is  $P(t,i)$ , when  $T(i, j) = \frac{P(t,i)}{P(t,j)} \geq 1$ , the data flow  $i$  will be given priority in scheduling, which will be described in detail in the scheduling.

(B) Formula for calculating the optimal route: After calculating the data stream weights in an accepted time, the optimal route selection for the entire line is required to calculate which is able to provide a reliable reference value for the scheduling algorithm and realize the rational use of resources. Suppose that the weight is  $P(t,i)$ , and the data stream type is  $T$ , instantaneous line current average utilization rate is  $R$  (taking the reality of network operating conditions into account), line standard bandwidth value is  $B$ , then the current bandwidth usage rate is  $B * R$  [6], the optimal route  $R_0$  is calculated as (4):

$$R_0 = \frac{L - (B * R) / 3}{\lambda_i} P(t,i) * T * 15\% \tag{4}$$

(the  $\lambda_i$  is line weight, the  $L$  is threshold line).

(C) An Optimal route flow diagram: To better illustrate the idea of optimal route selection, a flow chart below shows the optimal route in Fig. 3.

(D) Network congestion control processing: Of course, in actual operation, the network congestion is inevitable. Link traffic congestion can be defined to a period of time as the load is too large, resulting in link performance, reducing the data flow rate. In this paper, congestion can be described as the data flow, bandwidth, frequency and other parameters.

According to many references, we use the following model to process control network congestion:

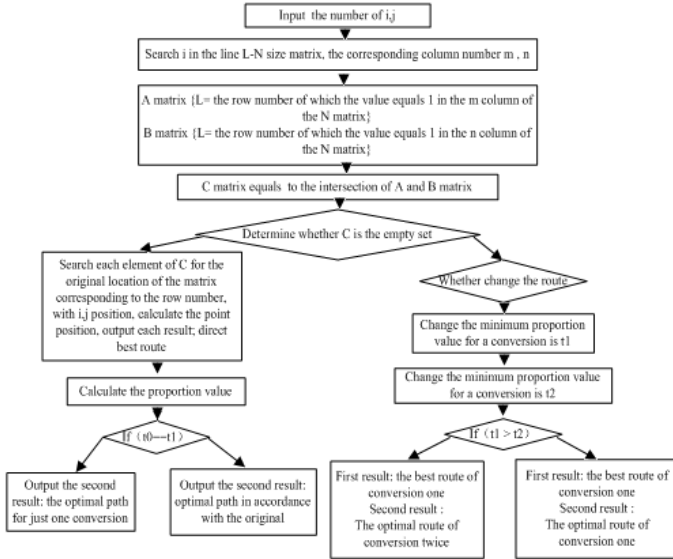


Fig. 3. Optimal route flow algorithm figure

$$\begin{cases} s = s_f * (1 - \frac{b}{b_{jam}}) \\ q = b * s \end{cases} \tag{5}$$

$s, s_f$  represent the data flow bandwidth and frequency on the current link ;  $b, b_{jam}$  is the current link bandwidth and the blocking density ;  $q$  represents the flow. Maximum flow conditions in different density :  $s_m = s_f / 2, b_m = b_{jam} / 2$  ; maximum flow:  $q_{max} = (s_m * s_m) / 2$  ( $s_m, b_m$  are constant and the constant density under the maximum flow rate). If  $b_m < b < b_{jam}, 0 < s < s_m, q < q_{max}$  , the network is in congestion state ; if  $b = b_{jam}, s = 0, q = 0$  , network congestion state occurs. In this scenario, this situation will be accepted by the above dynamic distribution of weight control, and progressively after influenza known weight value of each flow calculation; the scheme can largely reduce the appearance of congestion. Through the implementation of network monitoring link status, if there are a large number of data for transmission, we will forecast, timely and dynamically adjust the data on the network link, such as change data links and priorities class weights of the re-designation to ensure the smooth of the current link and data transmission quality themselves, under the links carrying capacity occurred in the conduct.

### 2.3 The Queue Scheduling Algorithm and Guardian

In this program, scheduling algorithm is mainly combined the idea of the expert system. And intelligent algorithms are using to determine the pre-order queue and automatic access. Through monitoring all network communication line and real-time

communication network line status, different queues will be into a different line by intelligent algorithms.

Guardian of the queue is simply like a watchdog. The queue dispatches and monitors operations of the queue. When the queue is in an abnormal situation, it will timely process, and storage the corresponding abnormal in abnormal knowledge base to facilitate the emergence of abnormal analysis.

With this scheduling algorithm, and guarding with the queue monitor, the knowledge base of intelligent algorithms can be increased through experiences running in a long term and the actual processing of data accumulated in the process. The perpetuation of this approach can achieve a more intelligent scheduling algorithm, and it is more able to increase the gold content to improve the network QoS.

### 3 Simulation Results

In this test, continuous high flow is sent through the network data in different formats which is an imitation of wide-area network data sent to test the performance on the end of admission control port. After a long period of stress tests, shown in Fig. 4, in a 10MB network resources link environment, the network average accepted the basic flow are small fluctuations up and down in 8MB. This result shows that the end of admission control side is able to obtain the change size of data flow between networks, the data is received at good quality, and the admission control can be achieved the rational allocation of network resources expected to ensure the network quality of service results.

At the same time, we have test the queue scheduling impacting on computer performance. Fig.5 is a state of the network monitoring program on the computer memory and CPU usage. According to the results shown in Fig. 5, for a long time to open the program, the memory usage maintains a relatively stable state, the CPU thread is small, and the impact on computer performance is also little, so that the process of communication across the network real-time monitoring network to ensure the normal operation of the computer program, in order to provide efficient QoS performance guarantees.

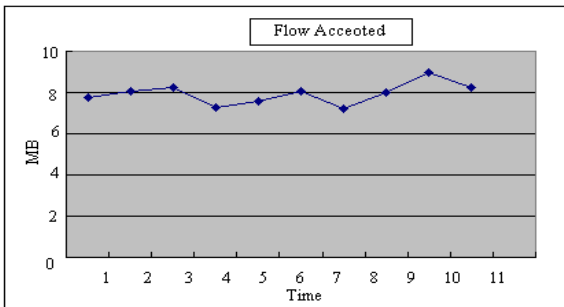
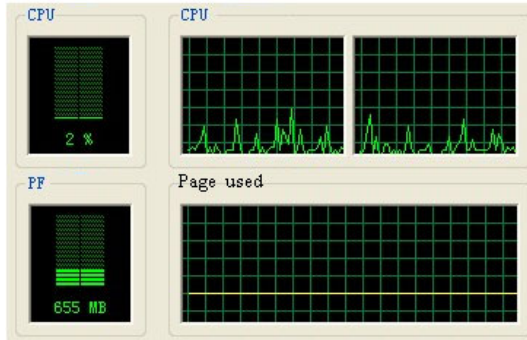


Fig. 4. The admission flow performance chart



**Fig. 5.** The performance of the computer diagram affected by the Scheduling algorithm

## 4 Conclusion

This paper proposed a network architecture for a new transport link to accept scheduling algorithm, by using the acceptance strategy which is integrated the global and partition weights calculated and the schedule which is as fair as possible to maximize the use of link resources effectively, to achieve the goal of avoiding network congestion as possible, obtaining the QoS guarantee and improvement,. The algorithm for further refinement is needed to be studied.

**Acknowledgments.** The Research is supported by the foundation of Nanjing University of Information Science & Technology (No.20070025), by The Natural Science Foundation of Higher Education of Jiangsu Province ,China( No. 11KJB510010), and by Priority Academic Program Development of Jiangsu Higer Education Institutions.

## References

1. Deng, X., Xing, C.: A QoS-oriented Optimization Model for Web Service Group. Dept. of Comput. Sci. & Technol., Tsinghua Univ, Beijing (2009); Computer and Information Science. In: Eighth IEEE/ACIS International Conference on ICIS 2009, June 1-3 (2009)
2. Lee, M., Kwon, B., Copeland, J.A.: Cross-Layer Design of DiffServ Architecture for End-to-End QoS in IEEE 802.16 Multi-Hop Mesh/Relay Networks with IEEE 802.11e WLANs. In: 2010 7th IEEE Consumer Communications and Networking Conference (CCNC), pp. 1–5 (2010)
3. Osali, S., Farahbakhsh, R., Abdolhosseini, A.M.: MEEAC: An enhanced scheme for supporting QoS granularity by multipath explicit endpoint admission control. In: 2010 International Conference on Telecommunications and Computer Networks (SoftCOM), pp. 55–59 (2010)

4. Liu, G.-Q., Zhu, Z.-L., Li, Y.-Q., Li, D.-C., Cui, J.-C.: Coll. of Software. Northeastern Univ., Shenyang (2009); A New Web Service Model Based on QoS. In: 2009 International Symposium on Intelligent Ubiquitous Computing and Education, May 15-16 (2009 )
5. Li, X., Chuah, E., Jo, Y.T., Kwong, H.G.: An optimal smooth QoS adaptation strategy for QoS differentiated scalable media streaming. In: 2008 IEEE International Conference on Inst. of High Performance Comput., Multimedia and Expo, Singapore, pp. 429–432 (2008)
6. Narsimha, G., Reddy, A.V., Kumar, B.S.: QOS Multicast Routing Based on Bandwidth and Resource Availability in Mobile Adhoc Networks. In: Seventh International Conference on Networking, ICN 2008, April 13-18. JNT Univ., Kakinada (2008)

# Template-Based Image Composition on the iPhone

Qingjie Sun, Wenjuan Chen, and Lei Liu

Animation School, Communication University of China, Beijing, China  
{sunqingjie,wjchen}@cuc.edu.cn, liulei224@163.com

**Abstract.** The Camera of mobile phone has become an important content source of photo-sharing social network site. In this paper, an image composition application on the iPhone mobile platform is presented. First, we design some templates of human body with various styles. Next, the user takes a photo or selects one from the photo library and composes it with one template via some gesture operations. Finally, the user can shake the iPhone, and then some animation effects may occur. The experimental results show the mobile application is full of fun.

**Keywords:** Image composition, iPhone, template, gesture.

## 1 Introduction

The iPhone mobile is one of the most popular smart phones in many countries now. One of the most important reasons is the apple's app store which has almost half million applications. Various applications make iPhone not only a phone but also a mobile platform.

An obvious difference between iPhone and personal computer is that iPhone has some sensors such as the proximity sensor, global positioning system, three-axis accelerometer, magnetometer, gyroscope. In this paper, we make use of the camera, multi-touch sensor and three-axis accelerometer to make image composition in a fun way.

In the remainder of this paper, we first introduce some related work in image composition and mobile application design. Next, our algorithms are discussed in detail. Then, experimental results are presented. Finally, conclusion is presented and some possible future works are discussed.

## 2 Related Work

Taking photo via digital camera or mobile has been an easy activity in our daily life. However, in many cases, the photo we take may not be what we want to express. In such case, we may combine parts of a set of photographs into a single composite picture which can convey our ideas. Segmenting a part from a source image and pasting it into the target image seamless are two key problems for most image composition tasks. D.Comaniciu [1] introduces a mean shift-based image segmentation algorithm which has been widely used in the computer vision

community. In [2], the interactive image segment is done by graph-cut optimization which has been used for a variety of tasks including stereo matching and optical flow. Alpha matte is used to separate transparent and partial objects in [3]. After segmented objects are obtained, alpha blending is usually used to insert them into the target image [4]. Chen [5] introduces a novel user interface for image composition based on sketch, and the source images are obtained from the internet. Liu [6] discusses the aesthetics properties of image composition and applies them to recompose images in order to increase the aesthetic appreciation of result. All these works aim to generate realistic image composition which seems no difference with photo taken by camera. In contrast, we try to make a mobile entertainment application full of fun.

Our work is also related to design and implementation of mobile applications. Gunnar [7] introduces an augmented reality application based on the iPhone platform. Emiliano [8] tries to quantitatively evaluate the performance of the iPhone's sensors, localization engine and networking stack while running a people-centric sensing application. In this paper, we make use of the iPhone's camera, multi-touch sensor and accelerometer to finish the composition in an intuitive and fun way.

### 3 Image Composition

In our application, two source images are provided to obtain the composition result. One source is the foreground layer, and the other is the background layer. We design some character's templates with cartoon style as the foreground layer whose face region is set as transparent part. The background layer is another character's photo taken by the user via mobile camera or selected from the photo library. We need to move, rotate and resize the background image so that its face region can align with the transparent region of the foreground layer. The composition is obtained by inserting the background's face region into the foreground layer.

The following figure illustrates the composition process. The left image in fig.1 is the foreground template with a transparent face region, the middle image in fig.1 is the background layer and the right image in fig.1 is the composition result.



Fig. 1. Composition process

#### 3.1 Interactive Image Transformation

The template is designed according to the iPhone's screen size, so the transparent region of foreground layer should not be changed. If we want to insert the face region

of the background into the foreground, we need to transform the face region of the background layer via moving, rotating and resizing operations. As the iPhone’s screen supports multi-touch function, we can define different gestures to interactively transform the background image.

Two kinds of gestures are supported in our application. One is one-finger drag gesture, and the other is two-finger pinch gesture. The simple one-finger drag gesture can be used to finish the moving operation, and its corresponding transformation is a translation matrix. The pinch gesture can be used to support the rotation and resizing operations, and its corresponding transformation is an affine matrix. The pinch gesture is a two-finger operation, and it has two forms: pinch open and pinch close, as illustrated in fig.2. The grey circles represent the initial states of the two fingers, and the dark circles are in the ending state of the two fingers. The state changes of the fingers can be revealed by a 3\*3 matrix.

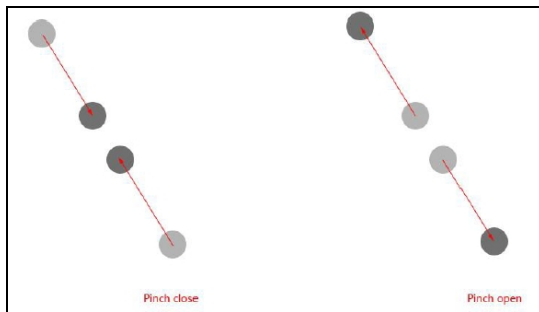


Fig. 2. Pinch gesture

When the touch event happens, we first detect the number of the touch points. If there is only one touch point, we can judge that it is a one-finger drag gesture, and we define the start touch point as  $P_b$ . If there are two touch points, we know it is a two-finger pinch gesture, and we define the start touch points as  $P1_b$  and  $P2_b$ .

When the touch continues, we will keep tracking the touch points and change the transformation matrix according to the gesture type and fingers state. For the drag gesture with one-finger touch, we denote the current touch point as  $P_c$ . If the user action belongs to the pinch gesture with two-finger touch, we denote the current two-finger touch points as  $P1_c$  and  $P2_c$ . The transformation matrix [9] can be determined by these touch points via the following equations.

When the touch belongs to one-finger gesture, the transformation matrix  $M$  can be expressed by the equation (1).

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ P_c.x - P_b.x & P_c.y - P_b.y & 1 \end{bmatrix} \tag{1}$$



When the touch belongs to two-finger gesture, we define the center of the current view of our mobile app as CV with the coordinate  $(C_x, C_y)$ .  $RC_1(x_1, y_1)$  is then defined as the coordinate of  $P1_b$  relative to the center point CV. At the same way,  $RC_2(x_2, y_2)$ ,  $RC_3(x_3, y_3)$  and  $RC_4(x_4, y_4)$  are defined as the coordinates of  $P2_b$ ,  $P1_c$  and  $P2_c$  relative to CV. To clearly express the transformation matrix in this case, we need to define some variables and equations.

$$D = (y_1 - y_2)^2 + (x_1 - x_2)^2 \quad (2)$$

$$a = (y_1 - y_2) * (y_3 - y_4) + (x_1 - x_2) * (x_3 - x_4) \quad (3)$$

$$b = (y_1 - y_2) * (y_3 - y_4) - (x_1 - x_2) * (x_3 - x_4) \quad (4)$$

$$\begin{aligned} tx = & (y_1 * x_2 - x_1 * y_2) * (y_4 - y_3) - \\ & (x_1 * x_2 + y_1 * y_2) * (x_3 + x_4) + \\ & x_3 * (x_2^2 + y_2^2) + x_4 * (x_1^2 + y_1^2) \end{aligned} \quad (5)$$

$$\begin{aligned} ty = & (x_1 * x_2 + y_1 * y_2) * (y_4 - y_3) + \\ & (y_1 * x_2 - x_1 * y_2) * (x_3 - x_4) + \\ & y_3 * (x_2^2 + y_2^2) + y_4 * (x_1^2 + y_1^2) \end{aligned} \quad (6)$$

$$M = \begin{bmatrix} a/D & -b/D & 0 \\ b/D & a/D & 0 \\ tx/D & ty/D & 1 \end{bmatrix} \quad (7)$$

As D plays as a denominator in equation (7), it can not be equal to zero. In fact, when D is close to zero, the affine transformation matrix M will degenerate into a translation matrix as equation (8). In our implementation, we set the minimum threshold of D as 0.1. If D is smaller than 0.1, we think it is close to zero.

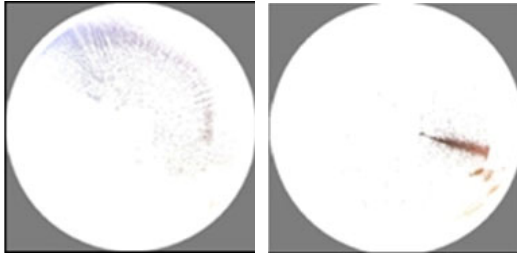
$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x_3 - x_1 & y_3 - y_1 & 1 \end{bmatrix} \quad (8)$$

### 3.2 Color Editing

If the lighting condition is not good, the background image we take via iPhone camera may be not satisfied and we may need to edit the color of its face region.

First, we convert the image from RGB space to HSV, and we can obtain the hue distribution illustrated in the left image of fig.3. For the Asian yellow skin, the hue falls into a 15 degree interval with the center at a fixed yellow position of the hue

wheel. The right image in fig.3 is obtained by scaling and rotating the hue distribution interval of the left image. For people outside Asia, we can also find a hue distribution interval, and the color editing can be done in the same way.



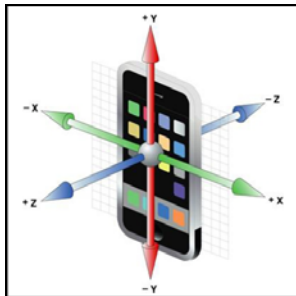
**Fig. 3.** Hue transformation

Color editing is an optional step. If the background image is taken under good conditions, it is not necessary to edit the image's color.

### 3.3 Interaction with Animation

In order to provide more fun to the application's users, we assign one or more animation clips with each template. We have designed animations with various styles, such as warm type, funny clip, even though some style with a little violence. We wish the animations can provide the users some pleasantly surprise feelings. After the image composition process is finished, the user can shake the iPhone to play one of the corresponding animation clips with a random way. To do this, we need to detect the shake action of the user via the accelerometer sensor.

There are three features for the shake action in our application. 1) Although the iPhone's accelerometer supports three-axis sensing, the X-axis is the main motion direction in our application. The coordinate system of accelerometer [10] is illustrated in fig.4. 2) Three or more peaks for shake should occur, and the forces with positive direction and negative direction should appear alternately. 3) The accelerometer data we obtain should fall in the interval from 0.75 to 1.25.



**Fig. 4.** Coordinate system of accelerometer

The accelerometer data is updated every 0.03 second in our application, and it can be considered as a signal [11]. The low frequency of this signal corresponds to the data exerted by the continuous force, such as gravity. The high frequency of this signal corresponds to the data exerted by the sudden force. Obviously, we wish to obtain the high-pass component of accelerometer data to capture the shake action. We denote the x-axis component of the accelerometer data as *acceleration.x*. The low-pass component of *acceleration.x* is denoted as *lowAcc*, and the high-pass of it is denoted as *highAcc*. They can be computed via the equation (9).

$$\begin{aligned} lowAcc &= acceleration.x * K + meanAcc * (1.0 - K) \\ highAcc &= acceleration.x - lowAcc \end{aligned} \quad (9)$$

In equation (9), *K* is the filtering factor which can reveal the role of the current acceleration plays, and *menaAcc* is the mean of *acceleration.x* in the previous time interval.

If the calculated accelerometer data meets the requirement mentioned above, a shake action is then detected, and one of the corresponding animation clips is played which is selected by the application randomly. The user can stop the animation clip, save the current animation frame and then share it with friends via email or facebook.

## 4 Results

We have implemented the application on the iPhone platform, and released it via the apple app store whose name is FaceFantasy. Fig.5 is the template with the Chinese traditional clothing styles, and fig.6 is the composition result by inserting a boy' face region into the template. Since the interaction via different gesture is a dynamic process, it is not suitable to be shown through static image. As a result, we present the composition result directly. At the top left corner of fig.6, there is dynamic visual mark which is used to remind the user about the shake action that he can do. Fig.7 is the final result which is obtained by overlaying one animation frame on the fig.6. The animation effect is optional, and the user can decide whether to shake the iPhone. The mobile application supports email and access to facebook and twitter. When the user obtains a satisfied composition, he can share it with friends if his iPhone supports internet access.



Fig. 5. Template image



Fig. 6. Composition



Fig. 7. Animation effect

## 5 Conclusion and Future Work

We discuss the design and implementation of our iPhone app named Face fantasy in the apple app store. The app allows the user to insert a face photo into a cartoon-style template to obtain a fun composition. After that, the user even can enjoy some animation effect by shaking the iPhone.

In the future, we will explore how to detect the face region according to skin color in order to simplify the interaction process. In addition, we will consider how to improve the visual effect of composition result based on some principles of color harmony.

**Acknowledgments.** We would like to thank Bin Zhu for insightful discussions, Xiaodan Liu for technical support, Jiaqing Si for providing photos and Peijia Hu for the art design.

The financial support of the 211project named “Intelligent Generation of Digital Media Content” from the Chinese ministry of education is gratefully acknowledged.

## References

1. Comanici, D., Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 603–609 (2002)
2. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics* 23(3), 309–314 (2004)
3. Wang, J., Cohen, M.F.: Simultaneous Matting and Compositing. In: *Proceedings of 24th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
4. Jia, J., Sun, J., Tang, C.-K., Shum, H.-Y.: Drag-and-Drop Pasting. *ACM Transaction on Graphics* 25(3), 631–636 (2006)
5. Chen, T., Cheng, M.-M., Tan, P., Ariel, S., Hu, S.-M.: Sketch2Photo: Internet Image Montage. *ACM Transactions on Graphics* 28(5), 1–10 (2009)
6. Liu, L., Chen, R., Lior, W., Daniel, C.-O.: Optimizing Photo Composition. *Computer Graphics Forum* 29(2), 469–478 (2010)

7. Gunnar, L.: Augmented Reality and Digital Genre Design—Situating Simulations on the iPhone. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality Arts, Media and Humanities, pp. 29–34 (2009)
8. Emiliano, M., Oakley, J.M.H., Lu, H., Lane, N.D., Petrsn, R.A., Campbell, A.T.: Evaluating the iPhone as a Mobile Platform for People-Centric Sensing Applications. In: Proceedings of International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems, pp. 41–45 (2008)
9. APPLE Inc. iPhone Application Programming Guide  
<http://www.andreasacca.com/public/downloads/documenti/iPhoneAppProgrammingGuide.pdf>
10. <http://iamkamty.com/blog/2011/01/iphone-tutorial-reading-the-accelerometer-switch-on-the-code/>
11. Allen, C., Appelcline, S.: iPhone in Action: Introduction to Web and SDK Development, 1st edn. Manning Publications (2009)

# Bag-of-Visual-Words Based Object Retrieval with E<sup>2</sup>LSH and Query Expansion

Zhao Yongwei, Li Bicheng, and Gao Haolin

National Digital Switching System Engineering Technology Center,  
Zhengzhou, Henan 450002, China

zhaoyongwei369@gmail.com, lbclm@163.com, holygao@126.com

**Abstract.** Boosted by the invention and wide popularity of SIFT image features and bag-of-visual-words (BoVW) image representation, object retrieval has progressed significantly in the past years and has already found deployment in real-life applications and products, but the traditional BoVW methods have several problems, such as: low time efficiency and large memory consumption, the synonymy and polysemy of visual words. In this article, a method based on E2LSH (Exact Euclidean Locality Sensitive Hashing) and query expansion is proposed. Firstly, E2LSH is used to hash local features of training dataset, and a group of scalable random visual vocabularies are constructed. Then, the visual vocabulary histograms and index files are created according to these random vocabularies. Finally, a query expansion strategy is used to accomplish object retrieval. Experimental results show that the accuracy of the novel method is substantially improved compared to the traditional methods, and it adapts large scale datasets well.

**Keywords:** Object retrieval, bag-of-visual-words method, E2LSH, query expansion.

## 1 Introduction

There are not only opportunities but also challenges to image retrieval along with the centralized and scaling up image data. Furthermore, we are always interested in the problem of specific object retrieval from an image database, in other words, given a query image in which a particular object has been selected, our system should return from its corpus a set of representative images in which that object appears. This is a harder problem than whole-image retrieval, since the object may be occluded, lit differently, or seen from different viewpoints in returned images. So, that is a very important issue to retrieve the specific object accurately and efficiently. Images are mostly characterized using the local features in object retrieval, such as SIFT 1 (Scale Invariant Feature Transform) and its expansion, each of which is invariant to image translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection, but it will lead to “curse of dimensionality” phenomenon because of their high dimension if we use the traditional index structure, such as R tree, K-D tree and the VA-File. The capabilities for near neighbor search of LSH2

(Locality Sensitive Hashing) and some other approximate query techniques are powerful, but the index is based on a single unit of SIFT feature, and all SIFT features should be matched one by one during object retrieval. So the time cost is greatly increased.

The Bag-of-Visual-Words method 345 was first proposed by Sivic, etc 3. Currently, due to its outstanding performance, it has become the mainstream method of object retrieval. While object retrieval based on the bag of visual words method is generally effective, there still exist several open issues. Firstly, there are problems of low time efficiency and large memory consumption in BoVW methods, which comes down to a great amount of high-dimensional nearest neighbor search tasks. For example, the traditional BoVW methods usually use K-Means to generate visual dictionary, that needs to assign a nearest center for each data point in every iteration of K-Means clustering and the vast majority of computation time is spent on calculating nearest neighbors between the points and cluster centers. Thus, as the visual data scales up, the time efficiency of BoVW decreases rapidly. The time complexity of K-Means is  $O(kNd)$ , where  $k$  is the number of cluster centers,  $N$  is the number of data points,  $d$  is the feature dimension. It was shown in 56 that such a time complexity is feasible for small values of  $k$ , but renders the algorithm intractable for large vocabularies ( $k > 10^5$ ). The time complexity is reduced to  $O(Nd \log k)$  by exploiting a nested structure of Voronoi cells, known as Hierarchical k-means (HKM) 5. In 6 it was shown that this reduced time complexity could also be achieved by a KD-forest approximation. The experiments of 6 demonstrated that vector quantization obtained by approximate k-means (AKM) is superior to HKM. Despite the use of inverted list structures to boost the search efficiency, memory storage still can be an issue for BoVW approaches. 6 reports the memory usage is 4.3GB for approximately 1.1M images. Besides, K-Means and its improvement algorithms 56 are not scalable 7, once new data points arrive, it needs to repeat the clustering procedure to obtain a new visual dictionary. That is obviously quite inconvenient.

Secondly, the traditional BoVW methods also have the problem of synonymy and polysemy of visual words, because K-Means and its improvement algorithms have the following defects: (1) They are variance-based, it awards more clusters to high frequency areas of the feature space, leaving less clusters for the remaining areas. Such attribute will lead to the problem of visual word synonymy—the phenomenon that several visual words depict the same kind of feature points. (2) They are non-robust. Points lying far from the center can significantly distort the position of the centre that they are assigned to, which will lead to the problem of visual word polysemy—the phenomenon that feature points depicted by the same visual word vary a lot. Therefore, a variety of attempts have been made to solve the two problems. Philbin and Chum, etc 8 proposed a “soft-assignment” method so that a high-dimensional descriptor would map to a weighted combination of visual words, rather than “hard-assigned” to a single word as in previous work. Wang and Li, etc 9 proposed a QP (quadratic programming) assignment method by using the linear construction weights of the neighboring visual words as the contribution functions.

The weights can be formulated and solved by the quadratic programming techniques. Jan et al. [10] used kernel functions to realize the soft mapping between local feature points and visual words. All these methods can in some degree solve the problems of visual word synonymy and polysemy of BoVW. However, no attempt has been made to resolve the problems of low time efficiency and large memory consumption mentioned above, since they are all based on K-Means and its improvement algorithms. Furthermore, Yang and Geng, et al. [11] presented a contextual object retrieval (COR) model that employed the visual context information together with the region of interest (ROI) to efficiently improve the distinguishability of objects in BoVW based image retrieval, Hsiao and Henry, et al. [12] put forward a topic-sensitive image retrieval method based on noise-proof relevance feedback to address the query drift problem in visual object retrieval.

$E^2$ LSH (Exact Euclidean LSH) [13] is a scheme of LSH (Locality Sensitive Hashing) [14] realized in Euclidean space. It is widely used in fast approximate nearest neighbor search for large scale high-dimensional data. The basic idea of  $E^2$ LSH is that several locality sensitive hashing functions are used to map high-dimensional data points into low-dimensional space ensuring the closer points in initial space still close to each other in low-dimensional space. Inspired by  $E^2$ LSH, we use  $E^2$ LSH to cluster local feature points of training image dataset to obtain a group of scalable randomized visual dictionaries, then, the visual vocabulary histogram and index files are constructed based upon these randomized visual dictionaries. Moreover, the tf-idf algorithm as proposed in text (and as applied to the BoVW) is applied to discount the influence of frequent discrete events, where an event is the occurrence of a visual word in an image. Besides, recent work [68] has shown that the traditional methods always suffer from poor recall: feature detectors often fail to fire even on near-duplicate images, and query regions often fail to contain the visual words needed to retrieve matches from the database. One very successful technique for boosting recall is query expansion [15], which achieves substantially better retrieval performance when the visual words in a query region are augmented using words taken from matching regions in the initial results set. Obviously, this method relies on sufficient recall from the initial query to get the process started, and can fail badly on queries with poor initial recall. Our approach based on  $E^2$ LSH specifically addresses the problem of recall from an initial query, and is therefore complementary to query expansion strategy.

The remainder of this paper is organized as follows. Section 2 briefly describes the  $E^2$ LSH algorithm. Section 3 presents the object retrieval framework based on  $E^2$ LSH and query expansion in detail. The experimental results are given in Section 4. And finally we conclude this paper in Section 5.



## 2 E<sup>2</sup>LSH

The key idea of E<sup>2</sup>LSH is to hash the points using several  $p$ -stable distribution based locality sensitive hashing functions, guaranteeing that points close to each other have a much higher probability of collision than points which are far apart. Locality sensitive hashing function  $h(\mathbf{v})$  and  $p$ -stable distribution are defined in 1314. In real applications, E<sup>2</sup>LSH usually concatenate several LSH functions by defining a family:

$$\mathcal{G} = \{g : S \rightarrow U^k\} \quad (1)$$

where  $g(\mathbf{v}) = (h_1(\mathbf{v}), \dots, h_k(\mathbf{v}))$ .

For each point  $\mathbf{v} \in \mathbb{R}^d$ , vector  $\mathbf{a} = (a_1, a_2, \dots, a_k)$  is obtained after mapping through  $g(\mathbf{v})$ . E<sup>2</sup>LSH utilizes the primary and the secondary hash function  $h_1$ ,  $h_2$  to hash the vector  $\mathbf{a}$  and establishes hash tables to store these data points. The two hash functions have the following form:

$$h_1(\mathbf{a}) = ((\sum_{i=1}^k r_i' a_i) \bmod \text{prime}) \bmod \text{size} \quad (2)$$

$$h_2(\mathbf{a}) = (\sum_{i=1}^k r_i'' a_i) \bmod \text{prime} \quad (3)$$

where  $r_i'$  and  $r_i''$  are random integers,  $\text{tablesize}$  is the size of the hash tables, and  $\text{prime}$  is a prime number, which equals  $2^{32} - 5$ .

## 3 Object Retrieval Based On E<sup>2</sup>LSH and Query Expansion

In this paper, we propose an object retrieval method based on E<sup>2</sup>LSH and query expansion as illustrated in Fig 1. Firstly, SIFT descriptors of image dataset and query object are extracted and mapped to the nearest neighbor visual words (hash bucket) using E<sup>2</sup>LSH to obtain frequency vectors. Secondly, the frequency vectors are weighted with tf-idf strategy to get the visual vocabulary histograms, which are stored on disk as index files to overcome the problem of large memory consumption. Then, the similarity matching between histograms of the query object and index files is made and an initial result is obtained. Finally, in order to improve retrieval performance the query region is enriched with additional information from the verified images of initial result and is retrieved again. We refer to the enriched region as a latent model of the object, which is a generalization of the idea of query expansion.

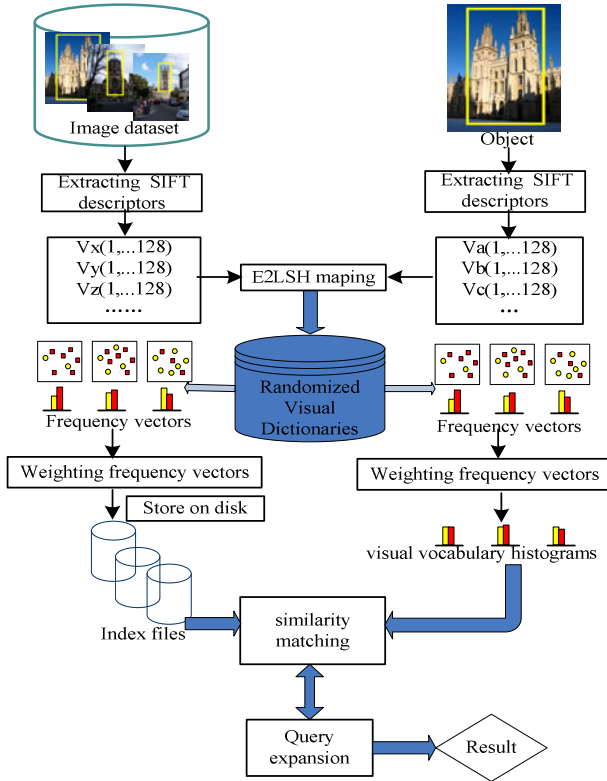


Fig. 1. The flow diagram of object retrieval based on E<sup>2</sup>LSH and query expansion

### 3.1 Randomized Visual Dictionaries Construction Based Upon E<sup>2</sup>LSH

Instead of using K-Means and its improvement algorithms, E<sup>2</sup>LSH is firstly employed to cluster local feature points of training dataset to obtain an ensemble of randomized visual dictionaries. The algorithmic flow is illustrated in Fig 2. The construction process of a single visual dictionary contains three steps as follows:

**Step 1: SIFT Detection.** The Oxford Buildings dataset 16 are selected as training dataset  $\mathcal{I} = \{I_1, I_2, \dots, I_k, \dots, I_{T-1}, I_T\}$ , then SIFT points of images in  $\mathcal{I}$  are detected, and a local feature pool  $\mathcal{R} = \{r_1, r_2, \dots, r_k, \dots, r_{K-1}, r_K\}$  is obtained, where  $r_k$  represents a 128-dimensional SIFT descriptor.

**Step 2: Visual Words Generation.** For an arbitrary SIFT point  $r \in \mathcal{R}$ , it is mapped through  $g_i$  to get a  $k$ -dimensional vector  $g_i(r)$ , the primary hash key  $h_1(g_i(r))$  and secondary hash key  $h_2(g_i(r))$  are calculated according to equations (2) and (3) respectively. SIFT points with the same primary and secondary hash key will be

stored in the same bucket, and the hash table  $T_i = \{b_1^{(i)}, b_2^{(i)}, \dots, b_k^{(i)}, \dots, b_{N_i-1}^{(i)}, b_{N_i}^{(i)}\}$  is obtained, where  $b_k^{(i)}$  represents an individual bucket of  $T_i$  and  $N_i$  is the number of buckets in  $T_i$ . Hash table  $T_i$  is regarded as a visual dictionary  $W_i$ , and each bucket of  $T_i$  is viewed as a visual word.

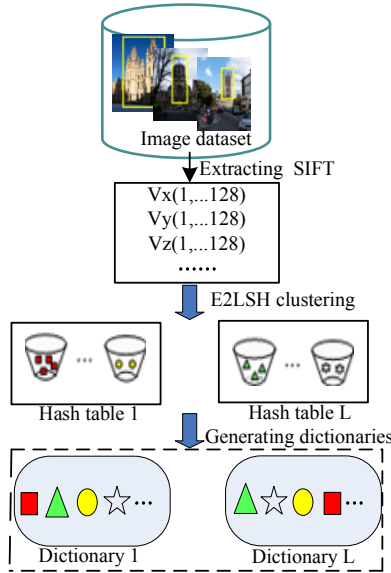


Fig. 2. The construction process of randomized visual dictionaries

**Step 3: Visual Words Filtering.** Generally, visual words with too high or too low frequency have weak discriminating power; these visual words can be abandoned to obtain a more compact dictionary  $W_i = \{w_1^{(i)}, w_2^{(i)}, \dots, w_k^{(i)}, \dots, w_{M-1}^{(i)}, w_M^{(i)}\}$ .

Obviously,  $L$  independent functions  $g_1, \dots, g_L$  can generate  $L$  randomized visual dictionaries according to the scheme above and the whole process is scalable. So, if a new image is added, we only need to detect its SIFT points and hash these points using  $E^2LSH$  to realize the dynamic expansion of randomized visual dictionaries.

To verify the effectiveness of  $E^2LSH$  clustering, we compare it to K-Means by clustering a same randomly generated dataset and the final results are shown in Fig 3. Comparing Fig 3 (a) with Fig 3 (b), we can find that K-Means has more cluster centers in dense areas and less cluster centers in sparse areas, while the distribution of  $E^2LSH$  bucket centers is more even.

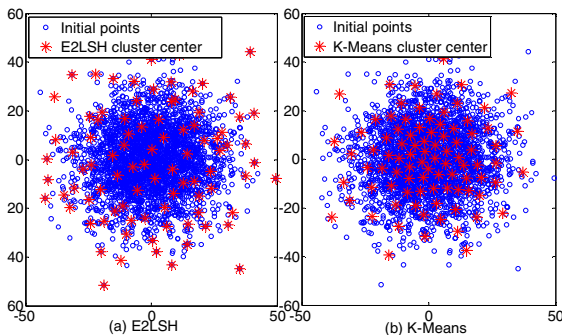


Fig. 3. Comparison of two data clustering algorithms

### 3.2 Visual Vocabulary Histograms Construction

Based on the randomized visual dictionaries generated above, visual vocabulary histograms can be extracted as the feature vectors of images. The distribution of visual words in an image can be estimated as follows.

For each visual word  $w$  in the visual dictionary  $W_i$ , define a function:

$$n_k(w) = \begin{cases} \omega(w, I), & \text{if } g(r_k) \in w \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where  $r_k$  is the SIFT point  $k$  of an image, and  $\omega(w, I)$  is the weight of visual word  $w$  in the image. To discount the influence of frequent discrete visual words in an image  $\omega(w, I)$  is calculated by tf-idf:

$$\omega(w, I) = \frac{tf(w, I) \times \log(N / n_i)}{\sqrt{\sum_{w \in I} [tf(w, I) \times \log(N / n_i)]^2}} \tag{5}$$

where  $tf(w, I)$  is the word frequency vector of  $w$  in image  $I$ ,  $N$  is the total number of training dataset,  $n_i$  is the number of images that contain visual word  $w$  in training dataset, the denominator is normalization factor. Then, we can count the distribution of visual word  $w$  in image  $I$ :

$$H(w) = \frac{1}{n} \sum_{k=1}^n n_k(w) \tag{6}$$

where  $n$  is the number of keypoint in image  $I$ . So, the visual vocabulary histogram generated by visual dictionary  $W_i$  is  $H_i(H_1, H_2 \dots H_M)$ .

### 3.3 Similarity Measurement and Query Expansion

Considering that the visual vocabulary histograms are space vectors, we will use the dot product to measure the similarity between images. Define the similarity measurement function:

$$Sim(Q, I) = \frac{1}{L} \sum_{i=1}^L \mathbf{H}_i(Q) \cdot \mathbf{H}_i(I) \quad (7)$$

where  $\mathbf{H}_i(Q)$  and  $\mathbf{H}_i(I)$  are visual vocabulary histograms that represent the query object and image  $I$  respectively.

In the text retrieval literature, a standard method for improving performance is query expansion, where a number of the highly ranked documents from the original query are reissued as a new query. This allows the retrieval system to use relevant terms without appearing in the original query. In 15, query expansion was brought into the visual domain. A strong spatial constraint between the query image and each result allows for an accurate verification of each return, suppressing the false positives which typically ruin text-based query expansion. These verified images can then be used to learn a latent feature model to enable controlled construction of expanded queries. The simplest well performing query expansion method is called average query expansion defined as follow:

$$d_{avg} = \frac{1}{N+1} \left( d_q + \sum_{i=1}^N d_i \right) \quad (8)$$

where  $d_q$  is the normalized tf vector of the query region, and  $d_i$  is the normalized tf vector of the  $i$ -th result. Then, a new query object model can be reconstructed by averaging a number of images descriptors, which are taken from the top verified results of the original result.

## 4 Experiments

### 4.1 Datasets and Evaluation Criteria

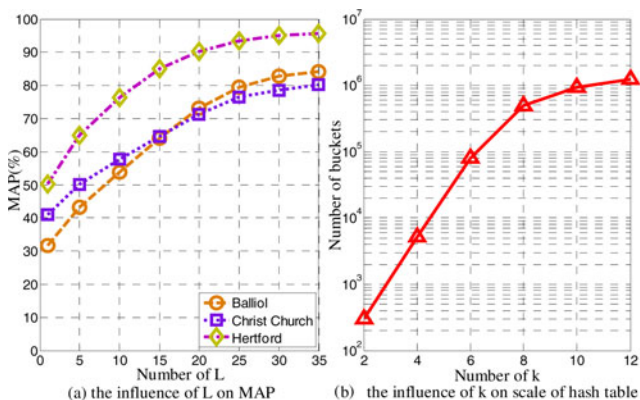
To evaluate the performance of our method, the Oxford Buildings dataset available from 16 is used. This is a relatively small set of 5K images with an extensive associated ground truth for 55 standard queries: 5 queries for each of 11 Oxford landmarks, such as Oxford Christ Church and Hertford. We also use an additional unlabeled dataset, Flickr1 16, which is assumed not to contain images of the ground truth landmarks. The images in this additional dataset are used as “distractors” for the system and provide an important test for the scalability of our approach. The two datasets are described in Tab 1.

**Table 1.** The number of descriptors for each dataset

Datasets	Number of images	Number of features
Oxford	5062	16,334,970
Flickr1	99782	277,770,833
Total	104,844	294,105,803

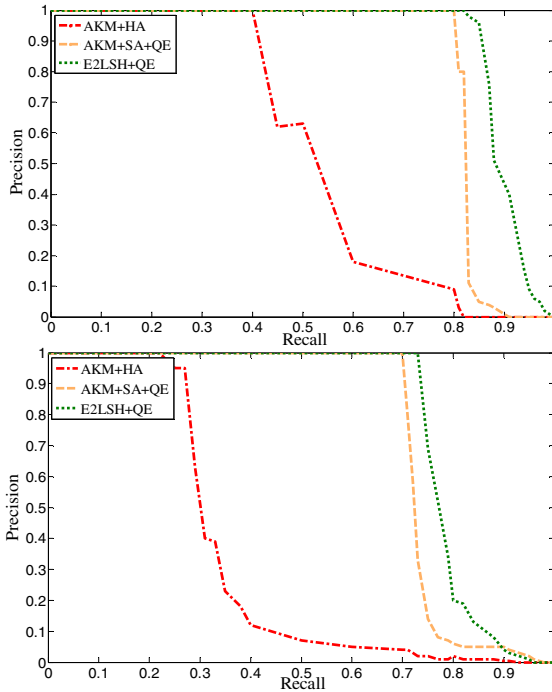
We use Average Precision (AP) computed as the area under the precision-recall curve. Precision is the number of retrieved positive images relative to the total number of images retrieved. Recall is the number of retrieved positive images relative to the total number of positives in the corpus. An ideal precision-recall curve has precision 1 over all recall levels, which corresponds to an Average Precision of 1. We compute an Average Precision score for each of the 5 queries for a landmark, and then average these to obtain a Mean Average Precision (MAP) for the landmark. For some experiments, in addition to the MAP, we also display precision-recall curves which can sometimes better illustrate the success of our system in improving recall.

In the experiments,  $L$  and  $k$  are two important parameters in E<sup>2</sup>LSH algorithm. A larger  $L$  will decrease the randomness of the algorithm, which is beneficial to the accuracy of the algorithm as shown in Fig 4 (a) ( the Balliol, Christ Church, and Hertford are used as the query object), but the efficiency of the algorithm will decrease as  $L$  increases.  $k$  has a large influence on the numbers of bucket in hash table (or the size of visual dictionary), as illustrated by Fig 4 (b). In general, the larger the size of visual dictionary is, the stronger its categorization ability is, but the efficiency of the algorithm will also decrease as  $k$  grows. So, comprehensively considering the efficiency and accuracy, we set  $L = 25$ ,  $k = 10$ .

**Fig. 4.** Influences of different E<sup>2</sup>LSH parameters

### 4.2 Results

First of all, we compare our approach ( $E^2$ LSH+QE) with the traditional hard assignment scheme in 6 (AKM+HA) and the frequently-used soft assignment scheme in 8 (AKM+SA+QE) to check the effectiveness for the synonymy and polysemy of visual words. The precision-recall curves for two queries are shown in Fig 5. It can be seen that our approach using the randomized visual dictionaries produces a few greater AP score than the method using the soft-assigned vocabularies with our implementation of approximate K-Means (AKM) of 6 when trained on the Oxford Buildings dataset. The results illustrate that our randomized visual dictionaries generated by  $E^2$ LSH are effective to overcome the synonymy and polysemy of visual words.



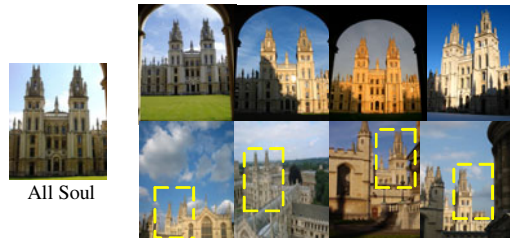
**Fig. 5.** The improvement in mAP obtained by using  $E^2$ LSH on two different queries (the above object is ashmolean\_3, the below is christ church\_5)

Tab 2 summarizes the results of different queries using average query expansion strategy, measuring their relative performance in terms of the MAP score. From the table, we can see that all the retrieval using average query expansion strategy perform much better than the original methods. It also proves that query expansion can boost the retrieval accuracy further upon  $E^2$ LSH.

**Table 2.** Summary of the relative performance of the different queries using average query expansion strategy

Query objects	AKM+SA (%)	AKM+SA+QE (%)	E <sup>2</sup> LSH (%)	E <sup>2</sup> LSH+QE (%)
Ashmolean	63.8	78.1	74.6	84.2
Bodleian	52.3	67.5	57.8	70.1
Christ Church	58.6	71.3	64.3	76.4
Hertford	79.8	87.3	88.6	94.8
Keble	75.3	90.3	87.8	89.1
Average	65.96	78.8	74.62	82.92

Fig 6 shows some example images returned by our method, the images of the above show results returned by the original query before query expansion, the images of the below show true positive images returned after query expansion. So, after query expansion, we can get many more examples of the object, some of which would be extremely challenging to the traditional methods, with, in some cases, very high levels of occlusion or large scale changes.

**Fig. 6.** Demonstrating the performance of our method on All Soul query

Besides, we evaluate the time efficiency of our method and the traditional methods. Firstly, extract about 1320000 SIFT points from 500 images which are clustered by approximate K-Means (AKM) of 6 and E<sup>2</sup>LSH. Then, the visual vocabulary histograms are constructed by the hard assignment method in 6 and our E<sup>2</sup>LSH method respectively. The results are shown in Fig 7. From Fig 7 (a), we can easily find that E<sup>2</sup>LSH used very fewer time than AKM under the same visual dictionary size, further more, the clustering time of AKM increases linearly as the keypoint data scales up, but the influence on E<sup>2</sup>LSH is minor. From Fig 7 (b), it can be seen that using hard assignment the average time for visual vocabulary histogram construction increases linearly as the visual dictionary scales up, however, the time of E<sup>2</sup>LSH is near steadiness.



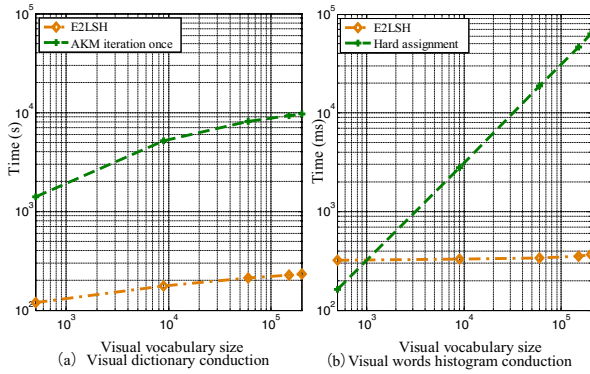


Fig. 7. Time efficiency comparison of different methods

## 5 Conclusion

In this article, instead of K-Means or its improvement algorithms we applied the E<sup>2</sup>LSH framework to local features clustering to generate the randomized visual dictionaries, which were used to construct the visual vocabulary histograms and index files. Then, in order to improve retrieval performance we therefore enriched the query region by query expansion strategy. Experimental results showed that our method can in some degree resolve the problems of visual word synonymy and polysemy and boost the retrieval accuracy substantially; besides, it adapts large scale datasets well. However, it should be pointed out that disk accesses and the query expansion scheme will increase some extra response time, what is a limitation towards much larger, “web-scale” image corpora. So, it is very necessary to study the more efficient retrieval methods.

**Acknowledgments.** This work is supported by grants from the Chinese National Natural Science Foundation under Contract. No. 60872142.

## References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
2. Slaney, M., Casey, M.: Locality-Sensitive Hashing for Finding Nearest Neighbors. *IEEE Signal Processing Magazine* 8(3), 128–131 (2008)
3. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos. In: *Proceedings of 9th IEEE International Conference on Computer Vision, Nice*, pp. 1470–1477 (2003)
4. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: *Proceedings of International Conference on Computer Vision, Beijing*, pp. 604–6 (2005)
5. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: *Proceedings of IEEE Conference on Computer Vision and Paaem Recognition, NewYork*, pp. 2161–2168 (2006)

6. Philbin, J., Chum, O., Isard, M., et al.: Object retrieval with large vocabularies and fast spatial matching. In: CVPR 2007: Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, June, 17-22, pp. 1-8 (2007)
7. Marée, R., Denis, P., Wehenkel, L., et al.: Incremental Indexing and Distributed Image Search using Shared Randomized Vocabularies. In: Proceedings of MIR 2010, Philadelphia, USA, pp. 91-100 (2010)
8. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 278-286 (June 2009)
9. Wang, J., Li, Y., Zhang, Y., et al.: Bag-of-Features Based Medical Image Retrieval via Multiple Assignment and Visual Words Weighting. In: IEEE Transactions on Medical Imaging (2011)
10. van Gemert, J.C., Snoek, C.G.M., Veenman, C.J., et al.: Comparing compact codebooks for visual categorization. *Computer Vision and Image Understanding* 114(4), 450-462 (2010)
11. Yang, L., Geng, B., Cai, Y., et al.: Object Retrieval using Visual Query Context. In: IEEE Conference on Computer Vision and Pattern Recognition, July 13, pp. 213-242 (2011)
12. Hsiao, J.-H., Chang, H.: Topic-Sensitive Interactive Image Object Retrieval with Noise-Proof Relevance Feedback. In: IEEE. ICASSP, pp. 869-873 (2011)
13. Datar, M., Immorlica, N., Indyk, P., et al.: Locality-sensitive hashing scheme based on p-stable distributions. In: SCG 2004 Proceedings of the Twentieth Annual Symposium on Computational Geometry, Brooklyn, New York, pp. 253-262 (2004)
14. Gionis, A., Indyk, P., Motwani, R.: Similarity search in high dimensions via hashing. In: Proceedings of the 25th International Conference on Very Large Data Bases (VLDB), Edinburgh, Scotland, pp. 518-529 (1999)
15. Sajina, R., Aghila, G., Saruladha, K.: A Survey of Semantic Similarity Methods for Ontology Based Information Retrieval. In: IEEE, Second International Conference on Machine Learning and Computing, pp. 297-302 (2010)
16. [http://www.robots.ox.ac.uk/\\_vgg/data/oxbuildings/](http://www.robots.ox.ac.uk/_vgg/data/oxbuildings/)
17. <http://www.flickr.com/>

# Online Learning Algorithm of Direct Support Vector Machine for Regression Based on Matrix Operation

Junfei Li and Xiaomin Chen

Center for Space Science and Applied Research Beijing, China  
ljf@cssar.ac.cn, chenxm@cssar.ac.cn

**Abstract.** With the development of support vector machine, its' application of online learning was paid close attention to more and more. In this paper ,we deduced an online learning algorithm for direct support vector machine (DSVM). By adding the bias square of hyperplane to the objective function, we introduced the mathematical mode of DSVM for regression which was extremely fast and simple, then we designed incremental and online learning algorithms by using the property of block matrix and kernel function matrix. Experimental results of simulation by using Mackey-Glass chaotic time series indicate the feasibility of the learning algorithm.

**Keywords:** Direct support vector machine, regression, online learning, matrix, kernel function.

## 1 Introduction

Because of its good performance, support vector machine[1] (SVM) has become a research hotspot in recent years. SVM was first proposed for pattern recognition problems, along with the insensitive loss function is introduced, SVM is applied to regression estimation and shows excellent learning performance. The standard SVM needs to solve the constrained quadratic programming problem, for regression, the scale of solving is doubled, the process of solving is time consuming. In order to simplify the original problem, a lot of scholars have studied from the angle of deformation. Reference [2] and [3] proposed a least squares support vector machine (LSSVM), which greatly reduce the complexity of computation. By thinking over the coefficient matrix of LSSVM is not positive definite and symmetric, a new regression direct support vector machine(DSVM) was put forward in [4]. Compared with least squares support vector machine, the new DSVM for regression only need solve the inverse matrix of a simple deformation on kernel matrix, the capacity of learning is similar to LSSVM, and the complexity of computation was reduced. In [5], model of regression least square support vector machine was analysed, incremental learning algorithm and on-line learning algorithm was designed that based on the block matrix calculation formula and the characteristics of kernel matrix. Based on [4] and [5], online learning algorithm of direct support vector machine for regression based on matrix operation was studied in this paper ,experimental results of simulation indicated the validity of the model.

The remaining of this paper is organized as follows: In section 2, the theory of DSVM for regression is reviewed, and online learning algorithm of direct support vector machine for regression based on matrix operation is deduced in the next section. Following section introduces the result of simulation. Section 5 summarizes the results and concludes our paper.

## 2 Regression Direct Support Vector Machine

According to [4], for the linear case, the model can be established through the deformation of LSSVMR that the bias square of hyperplane is added to the objective function:

$$\min_{w,b,\xi} \frac{1}{2}(w^T + w + b^2) + \frac{C}{2} \sum_{i=1}^n \xi_i^2 \tag{1}$$

$$s.t. w^T x_i + b + \xi_i = y_i, i = 1, \dots, n$$

$$L(w, b, \xi, \alpha) = \frac{1}{2}(w^T w + b^2) + \tag{2}$$

$$\frac{C}{2} \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (w^T x_i + b + \xi_i - y_i)$$

$$\begin{cases} \frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i x_i = 0 \\ \frac{\partial L}{\partial b} = b - \sum_{i=1}^n \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = C \xi_i - \alpha_i = 0 \\ \frac{\partial L}{\partial \alpha_i} = w^T x_i + b + \xi_i - y_i = 0 \end{cases} \tag{3}$$

The Lagrangain function of (1) is as equation (2).

By using Karush-Kuhn-Tucker (KKT) conditions, the partial derivative of equation (3) will be 0.

Mark all of the input samples with matrix  $A_{n \times m}$ , every row of  $A$  is a sample  $x_i$ , we can get the results from (3):

$$w = A^T \alpha, \quad b = e^T \alpha, \quad \xi = C^{-1} \alpha,$$

Combining the substitution with these for the last equation of (3):

$$AA^T \alpha + ee^T \alpha + C^{-1} \alpha = Y$$

So Lagrangain multipliers can be obtained as follow:

$$\alpha = (AA^T + E + C^{-1}I)^{-1}Y \tag{4}$$

( $E = ee^T$  is a square matrix with all of its elements being 1)

Linear regression function for DSVM is as follow:

$$f(x) = w^T x + b = (Ax + e)^T \alpha \tag{5}$$

For nonlinear regression, the samples set  $A$  is mapped into the feature space by kernel function as to  $\phi(A)$ , the problem is described as follows:

$$\begin{aligned} \min_{w,b,\xi} & \frac{1}{2}(w^T w + b^2) + \frac{C}{2} \sum_{i=1}^n \xi_i^2 \\ \text{s.t.} & w^T \phi(x_i) + b + \xi_i = y_i, i = 1, \dots, n \end{aligned} \tag{6}$$

Similarly, by solving the Lagrangain function of (6) and using the KKT conditions, we can get :

$$w = (\phi(A))^T \alpha, b = e^T \alpha, \xi = C^{-1} \alpha$$

Further:

$$\phi(A)\phi(A)^T \alpha + ee^T \alpha + C^{-1} \alpha = Y$$

By using kernel matrix:

$$Q = Q(A, A^T) = (\phi(A), \phi(A)^T)$$

The result will be:

$$Q\alpha + ee^T \alpha + C^{-1} \alpha = Y$$

So Lagrangain multipliers can be obtained as follow:

$$\alpha = (Q + E + C^{-1}I)^{-1}Y \tag{7}$$

Nonlinear regression function for DSVM is as follow:

$$f(x) = w^T \phi(x) + b = (Q(A, x) + e)^T \alpha \tag{8}$$

For the nonlinear case, when the kernel function is the linear kernel, namely  $Q = AA^T$ , which becomes the linear case, namely formula (7) and (8) are respectively transformed into formula (4) and (5). So the linear and nonlinear case can be unified.

We make the substitution:

$$H = Q + E + C^{-1}I$$

Because the kernel function satisfies the Mercer condition,  $Q$  is symmetric and positive semidefinite. Apparently  $E$  is also symmetric and positive semidefinite,

$C^{-1}I$  is symmetric and positive definite. So  $H$  is symmetric and positive definite matrix which ensures the solutions for the equations is unique.

### 3 Online Learning Algorithm of Direct Support Vector Machine for Regression Based on Matrix Operation

For realizing online learning, we need study the incremental learning. The fast incremental algorithm that based on inverse matrix is studied first, then follows the online learning algorithms that removes the oldest sample according to the concept of time window.

#### 3.1 Incremental Learning Algorithm

Suppose a given samples at time  $t$  is  $\{x(t), y(t)\}$ , where  $x(t) = [x_1, x_2, \dots, x_t]$ ,  $y(t) = (y_1, y_2, \dots, y_t)^T$ ,  $x_i \in R^m$ ,  $y_i \in R$ . Similarly, the kernel function matrix  $Q$ , Lagrange multipliers  $\alpha$  and  $H$  are all functions about time  $t$ :

$$Q_t(i, j) = k(x_i, x_j), \quad i, j = 1, 2, \dots, t$$

$$\alpha(t) = (\alpha_1, \alpha_2, \dots, \alpha_t)^T$$

$$H(t) = Q_t + E + C^{-1}I$$

The equation (8) yields the following modification:

$$y(x, t) = (Q_t(A, x) + e)^T \alpha(t) \tag{9}$$

In order to achieve incremental learning, we let  $U(t) = H(t)^{-1}$ . Then we will use the skills of matrix calculation solving  $U(t+1)$ .

At time  $t$ , the kernel matrix  $Q_t$  is a square matrix:

$$Q_t = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_t) \\ \vdots & \ddots & \vdots \\ k(x_t, x_1) & \dots & k(x_t, x_t) \end{bmatrix}$$

So,  $H(t)$  is a square matrix of  $t \times t$ :

$$H(t) = \begin{bmatrix} k(x_1, x_1) + 1 + 1/C & \dots & k(x_1, x_t) + 1 \\ \vdots & \ddots & \vdots \\ k(x_t, x_1) + 1 & \dots & k(x_t, x_t) + 1 + 1/C \end{bmatrix}$$



$$\begin{bmatrix} A_{11} & E \\ E^T & d \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + rr^T z \tag{11}$$

Among (11)

$$r = (E^T A_{11}^{-1}, -1)^T$$

$$z = (d - E^T A_{11}^{-1} E)^{-1}$$

Use (11), we can obtain  $U(t+1)$  :

$$U(t+1) = H(t+1)^{-1} = \begin{bmatrix} H(t) & V(t+1) \\ V(t+1)^T & v(t+1) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} H(t)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + r(t+1)r(t+1)^T z(t+1) \tag{12}$$

$$= \begin{bmatrix} U(t) & 0 \\ 0 & 0 \end{bmatrix} + r(t+1)r(t+1)^T z(t+1)$$

Among (12)

$$r(t+1) = (V(t+1)^T U(t), -1)^T$$

$$z(t+1) = (v(t+1) - V(t+1)^T U(t) V(t+1))^{-1}$$

As can be seen from (12),  $U(t+1)$  can be solved from  $U(t)$  through recursive step which avoid inverse operation of large matrix and improve the computational efficiency greatly.

### 3.2 Online Learning Algorithm Based on Time Window

Suppose the length of time window is  $l$ , so the number of samples at time  $t$  is  $l$ ,  $H(t)$  is divided for blocks anew:

$$H(t) = \begin{bmatrix} f(t) & F(t)^T \\ F(t) & W(t) \end{bmatrix}$$

The block elements of  $H(t)$  is as follow:

$$f(t) = k(x_1, x_1) + 1 + 1/C$$

$$F(t) = [k(x_1, x_2) + 1, \dots, (k(x_1, x_l) + 1)]^T$$



$$W(t) = \begin{bmatrix} k(x_2, x_2) + 1 + 1/C & \cdots & k(x_2, x_l) + 1 \\ \vdots & \ddots & \vdots \\ k(x_l, x_2) + 1 & \cdots & k(x_l, x_l) + 1 + 1/C \end{bmatrix}$$

Where  $f(t)$  is a scalar quantity,  $F(t)$  is a column vector with dimensionality of  $(l-1)$ ,  $W(t)$  is a square matrix of  $(l-1)(l-1)$ .

At time  $t+1$ , a new sample  $(x_{l+1}, y_{l+1})$  is appended, the oldest sample  $(x_1, y_1)$  is discarded, so  $H(t+1)$  is also a square matrix of  $l \times l$  and can be divided for blocks as follow:

$$\begin{aligned} H(t+1) &= \begin{bmatrix} k(x_2, x_2) + 1 + 1/C & \cdots & k(x_2, x_{l+1}) + 1 \\ \vdots & \ddots & \vdots \\ k(x_{l+1}, x_2) + 1 & \cdots & k(x_{l+1}, x_{l+1}) + 1 + 1/C \end{bmatrix} \\ &= \begin{bmatrix} W(t) & V(t+1) \\ V(t+1)^T & v(t+1) \end{bmatrix} \end{aligned}$$

Among the above block matrix:

$$v(t+1) = k(x_{l+1}, x_{l+1}) + 1 + 1/C$$

$$V(t+1) = [k(x_{l+1}, x_2) + 1, \dots, k(x_{l+1}, x_l) + 1]^T$$

Using the skills of solving matrix in (12), we can solve  $U(t), U(t+1)$ :

$$\begin{aligned} U(t) &= H(t)^{-1} = \begin{bmatrix} f(t) & F(t)^T \\ F(t) & W(t) \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & W(t)^{-1} \end{bmatrix} + r_1(t)r_1(t)^T z_1(t) \end{aligned} \tag{13}$$

$$\begin{aligned} U(t+1) &= H(t+1)^{-1} = \begin{bmatrix} W(t) & V(t+1) \\ V(t+1)^T & v(t+1) \end{bmatrix}^{-1} \\ &= \begin{bmatrix} W(t)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + r_2(t+1)r_2(t+1)^T z_2(t+1) \end{aligned} \tag{14}$$

Note that, among (13)

$$r_1(t) = (-1, F(t)^T W(t)^{-1})^T,$$

$$z_1(t) = (f(t) - F(t)^T W(t)^{-1} F(t))^{-1}$$

And among (14)

$$r_2(t+1) = (V(t+1)^T W(t)^{-1}, -1)^T,$$

$$z_2(t+1) = (v(t+1) - V(t+1)^T W(t)^{-1} V(t+1))^{-1}$$

Where  $f(t) \in R$  and  $F(t) \in R^{l-1}$  is the element and vector which be formed due to removing the oldest sample because of samples' migration in time window, and  $v(t+1) \in R$  and  $V(t+1) \in R^{l-1}$  is alike but owe to updated sample.

So far, for realizing online learning algorithm of direct support vector machine for regression based on matrix operation, there is only a problem remained that is how to solve  $W(t)^{-1}$  from  $U(t)$ . We rewrite  $U(t)$  as follow:

$$U(t) = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & W(t)^{-1} \end{bmatrix} + \begin{bmatrix} z_1(t) & z_1(t)(-F(t)^T W(t)^{-1}) \\ z_1(t)(-W(t)^{-1} F(t)) & z_1(t)(W(t)^{-1} F(t) F(t)^T W(t)^{-1}) \end{bmatrix} \tag{15}$$

Looking at (15), we can see if we know  $U(t)$  already, according to  $U_{11}$ ,  $z_1(t)$  can be solved.

Again, according to  $U_{12}$ ,  $z_1(t)$ ,  $(-F(t)^T W(t)^{-1})$  can be solved.

Similarly, according to  $U_{21}$ ,  $z_1(t)$ ,  $(-W(t)^{-1} F(t))$  can be solved.

Last, according to  $U_{22}$ ,  $z_1(t)$ ,  $(-F(t)^T W(t)^{-1})$  and  $(-W(t)^{-1} F(t), W(t)^{-1})$  can be solved.

When  $W(t)^{-1}$  is solved, according to (14), the recursive algorithm solving  $U(t+1)$  from  $U(t)$  is accomplished. The advantage of the recursive algorithm is to decrease the computational complexity and accelerate the speed of solving large inverse matrix.

Conclude all of that what were discussed above, the integrated algorithm of online learning algorithm of direct support vector machine for regression based on matrix operation is as follows:

- Step 1: initializing  $H(t), U(t)$ ;
- Step 2: according to  $U(t)$ , solving  $W(t)^{-1}$ ;
- Step 3: according to  $U(t)$ , solving  $\alpha(t)$ ;

Step 4: sampling new data  $(x(t+1), y(t+1))$  , according to  $U(t)$  and  $x(t+1)$  predicting  $\overline{y(t+1)}$  which can be compared with  $y(t+1)$  for evaluating the algorithm.

Step 5: solving  $U(t+1)$  , time window moving, let  $U(t) = U(t+1)$  ,goto Step 2.

### 4 The Result of Simulation

In order to verify the reliability of the algorithm, we have used Mackey-Glass chaotic time series for simulation. The samples are generated by Mackey-Glass chaotic time series which is formulized by delay differential equation as follow:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+[x(t-\tau)]^{10}} - 0.1x(t) \tag{16}$$

Where  $\tau$  is a delay parameter which must be greater than 16.8 for generating chaotic time series, we let  $\tau = 30$  in our simulation operation.

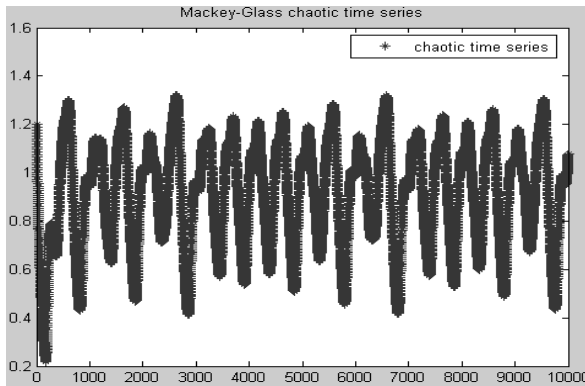


Fig. 1. Mackey-Glass chaotic time series

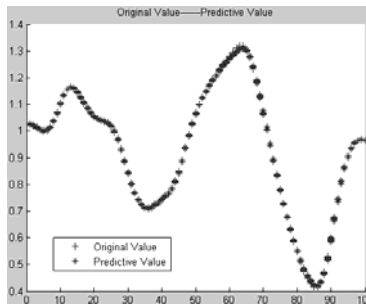


Fig. 2. Simulation results of Mackey-Glass chaotic time series

Gaussian kernel function which is defined as  $K(x, x_i) = e^{-\frac{1}{\gamma} \|x - x_i\|^2}$  is adopted in our simulation, when  $\gamma = 1.2$  and regularization parameter  $C = 300$ , mean square error is as follow:

$$\left(\frac{1}{l} \sum_{i=1}^l (\bar{y}_i(t+1) - y_i(t+1))^2\right)^{\frac{1}{2}} = 0.003 (l = 100).$$

The picture of time series generated is as figure 1.

The simulation results are shown in Figure 2. From it, we can see that the predictive values keep a good consistency with original value.

## 5 Conclusions

The paper introduces the mathematical model of direct support vector machine for regression first, and then online learning algorithm of direct support vector machine for regression based on matrix operation is put forward. Simulation results indicate that the algorithm is of good ability for regression of online learning.

**Acknowledgments.** The work was done while Li Junfei was with Laboratory of Space Integrated Electronics Technology Research at Center for Space Science and Applied Research.

## References

1. Vapnik, V.N.: The nature of statistical theory (1995)
2. Suykens, J.A.K., Vandewalle, J.: Least square support vector machine classifiers. *Neural Processing Letters*, 293–300 (1999)
3. Suykens, J.A.K., Vandewalle, J.: Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems*, 1109–1114 (2000)
4. Du, Z., Liu, S.-Y., Wu, Q.: Direct support vector machine for regression. *Systems Engineering and Electronic*, 178–181 (2009)
5. Zhang, H., Wang, X.: Incremental and Online Learning Algorithm for Regression Least Squares Support Vector Machine. *Chinese Journal of Computers*, 400–406 (2006)
6. Zhang, J.: Linear model's parameters estimation method and its improvement (1992)

# Cable Geometric Modeling Technology in Virtual Environment

Ma Long, Liu Pengyuan, and Mi Shuangshan

Ordinance Engineering College, Shijiazhuang, China

Malong198809@163.com, Lpy\_jx@sina.com, mishsh@sina.com

**Abstract.** Cable is flexible object. Its deformation is defined by material properties, cross-section sizes and length conservation gravity and friction. So it consists of variable shapes. Cubic uniform B-Spline function is chosen as the mathematical model of cable centerline in this paper. Energy-Optimization method is used to search for the positions of the control points. Firstly, a suspending cable model with both ends fixed is established; Based on that, other cable models with the constraints of endpoint tangents or passing through given points are built up by Lagrange Multiplier method; Then the method named Same-Proportion Dichotomy is used to cope with the constraint of length conservation. Secondly, two kinds of cable models are established. One is with one end hanging in the air established by Gradient-Descent method, the other is with one end dragging on the ground by Gradient Projection Descent method. Because traditional virtual modeling technology mostly focused on rigid parts or small deformation components, these studies consist of important meaning for the farther development of virtual modeling technology.

**Keywords:** Cable Geometric Modeling, Energy-Optimization Method, B-Spline, Gradient Projection Descent Method.

## 1 Introduction

The development of virtual maintenance technology has effectively improved the efficiency of maintenance training[1]. But the current studies mostly focus on rigid parts or small deformation components, research on large deformation bodies such as cable is still in the initial stage. Some virtual reality softwares, such as *veo*, have realized collision detection[2], interactive contact simulation as well as multibody real-time dynamics[3]. However, because of no feasible ways to deal with flexible bodies, deformations occurring in the physical world cannot be simulated.

This paper mainly concerns about homogeneous medium cables whose length is much bigger than cross-section sizes. Nowadays several different modeling methods have been proposed. Multipoint line is used to describe the cable centerline[4], which realized cable assembly and path planning. It is convenient to control the shape and length, but the fidelity is poor. R. C. Veltkamp and W. Wesselink adopts finite element method to establish cable model[5]. Since the complexity of path planning increases exponentially with the number of degrees of freedom that an extensive

exploration of the configuration space is very hard[6]. Cui Tong *et al.* use a method based on spring-mass system[7], it is difficult to realize maintenance process simulation as the calculation amount is so large.

With consideration of material properties, cross-section sizes, constraints of endpoint tangents, passing of given points, length conservation, gravity and friction, this paper adopts B-Spline function to describe cable centerline and Energy-Optimization method to search for the control points. Because of small amount of calculation and high fidelity, the model can be used in virtual maintenance process simulation.

## 2 The Choice of Spline Function

In geometry, cable is a curve with certain section. So the core of building the mathematical model is choice of proper mathematical function to describe cable centerline. Cubic uniform B-Spline function has the characteristics of simply modeling and small scale of calculation, applied to the circumstance that with little length change of the characteristics polygon, in this research monoclonal-segmental quasi-uniform cubic B-Spline is introduced to describe the cable. Set the node vector as (0,0,0,0,1,1,1,1), then its mathematical function is:

$$W(u) = \sum_{i=0}^3 V_i B_{i,3}(u) \tag{1}$$

$$\begin{cases} B_{0,3} = -u^3 + 3u^2 - 3u + 1 \\ B_{1,3} = 3u^3 - 6u^2 + 3u \\ B_{2,3} = -3u^3 + 3u^2 \\ B_{3,3} = u^3 \end{cases} \quad 0 \leq u \leq 1 \tag{2}$$

$V_i(i = 0,1,2,3)$  is the control point of the curve, and  $u$  is a curve parameter. Monoclonal-segmental quasi-uniform cubic B-Spline refers to the curve with 3 power-times, four control points, and the first and last control points coincided with the two end points[8]. Because of small amount of calculation, smooth appearance and flexible control of deformation, it is suitable to describe the cable centerline.

## 3 Suspending Cable with Both Ends Fixed

Energy-Optimization Method is proposed to search for the positions of the four control points in this paper, then the cable centerline can be drawn out.

### 3.1 Establishment of Mathematical Model

Take gravity into consideration, by Gossard[9] deformation energy equation, the cable deformation energy equation is

$$E = \int_0^1 (\alpha W_u(u)^2 + \beta W_{uu}(u)^2) du - 2 \int_0^1 W(u)g(u)du \tag{3}$$

$\alpha$  is bending stiffness parameter,  $\beta$  is stretching stiffness parameter, and  $g(u)$  is the unit length weight.  $W_u(u)$  and  $W_{uu}(u)$  is the first and second derivative vector of the parameter equation  $W(u)$  respectively.

By equation (1), (2):

$$E = \sum_{i=0}^3 \sum_{j=0}^3 (V_i * V_j * S_{i,j}) - 2 \sum_{i=0}^3 V_i * T_i \tag{4}$$

$$S_{i,j} = \int_0^1 (\alpha B_{i,3}'(u) B_{j,3}'(u) + \beta B_{i,3}''(u) B_{j,3}''(u)) du \tag{5}$$

$$T_i = \int_0^1 B_{i,3}(u) g(u) du \tag{6}$$

For a both ends fixed cable, the constraints can be expressed as:

$$\begin{aligned} V_0 &= W(0) = P_0 \\ V_3 &= W(1) = P_1 \end{aligned} \tag{7}$$

Without other constraints, get the minimum of  $E$  by designing  $V_1, V_2$ :

$$\begin{cases} \frac{\partial E}{\partial V_1} = 0 \\ \frac{\partial E}{\partial V_2} = 0 \end{cases} \tag{8}$$

With major variable elimination technique

$$V_i = -(\sum_{j=0, j \neq i}^3 V_j * S_{i,j} - 2 * T_i) / (2 * S_{i,i}) \tag{9}$$

Specific steps for solution are as follows:

- (1) Linear interpolate  $V_1, V_2$  between  $V_0$  and  $V_3$ ;
  - (2) Calculate the values of  $S_{i,j}$  ( $i, j = 0,1,2,3$ ) and  $T_i$  ( $i = 1,2$ ) by equation (2), (3), (4), (10) and (11);
  - (3) Give the maximum of iterating times, and perform iteration: Calculate  $V_1, V_2$  by equation (9);
  - (4) Use the control points obtained and B-Spline function drawing cable centerline.
- Since the methods above do not take friction and length conservation into account, they are available only for suspending cables with both ends fixed. See Figure 1.

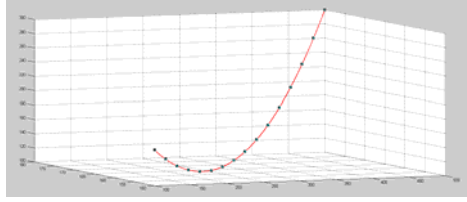


Fig. 1. The model of suspending cable centerline with both ends fixed

### 3.2 Mathematical Model of Constraints

Deformation of a suspending cable with both ends fixed is often affected by constraints of tangents, passing through given points and length conservation. The former two can be processed with Lagrange Multiplier method [2], and for the last one, a method named Same-Proportion Dichotomy is proposed.

Suppose the given length of a cable is  $L_0$ . For a monoclonal-segmental quasi-uniform cubic B-Spline curve, its length increases when the control points  $V_1$  and  $V_2$  moves away from the line  $V_0V_3$  and decreases when  $V_1$  and  $V_2$  close to  $V_0V_3$ . So the following algorithm is effective in adjusting length of the cable.

1) Get the length of the suspending cable:

$$L = \int_0^1 \sqrt{W_x'^2(u) + W_y'^2(u) + W_z'^2(u)} du \tag{10}$$

$W_x'(u)$ ,  $W_y'(u)$ ,  $W_z'(u)$  is the first derivative of  $W_x(u)$ ,  $W_y(u)$ ,  $W_z(u)$ , which is component of  $W(u)$  along corresponding axis respectively in 3-d orthogonal coordinate system.

2) Use Same Proportion Dichotomy for iteration:

Initialize  $V_i^{(0)} = V_i'$ ,  $V_i^{(1)} = V_i$  ( $V_i'$  is trisection of the line segment  $V_0V_3$ ,  $i = 1,2$ ).

For  $L^{(k)} > L_0$ , if  $L^{(k)} - L_0$  consists of a different sign from  $L^{(k-1)} - L_0$ , set

$$V_i^{(k+1)} = \frac{1}{2}(V_i^{(k)} + V_i^{(k-1)}) \tag{11}$$

if  $L^{(k)} - L_0$  consists of a same sign with  $L^{(k-1)} - L_0$ , set

$$V_i^{(k+1)} = V_i^{(k)} + \frac{1}{2}(V_i^{(k)} - V_i^{(k-1)}) \tag{12}$$

Where  $V_i^{(k)}$ ,  $L^{(k)}$  is coordinate of  $V_i$  and length of the cable after k times iteration.

For  $L^{(k)} < L_0$ , set

$$\overrightarrow{V_i^{(0)}V_i^{(k+1)}} = 2 * \overrightarrow{V_i^{(0)}V_i^{(k)}} \tag{13}$$

Repeat 1), 2) until meet the stop condition.



3) Stop condition:  $L^{(k)} - L_0 \leq \delta_1 L_0$ , where  $\delta_1$  is positive value much smaller less than 1.

## 4 The Cable with One Free End

### 4.1 Free End Suspending

In virtual maintenance environment, a cable with one end ( $V_3$ ) fixed and the other ( $V_0$ ) free is often encountered. According to mathematical optimum theory, negative gradient is the steepest drop direction of a function [10]. Therefore Gradient Descent method is adopted to search position of the end  $V_0$ . Specific procedures are as follows:

1) Initialize the starting points of Gradient Descent method. Reset coordinate system, make the projection of  $V_3$  on the ground as the origin and  $xz$  coordinate surface as the horizontal plane. Initialize coordinate of  $V_0$  as  $V_{0,x}^{(0)} = 0$ ,  $V_{0,y}^{(0)} = h - L_0$ ,  $V_{0,z}^{(0)} = 0$ .

2) Take  $V_0$  and  $V_3$  as the first and the fourth control point, take constraints of fixed end, tangent and passing of given points into consideration, adopt Energy-Optimized Method searching for the position of  $V_1$  and  $V_2$ , and calculate the length of cable ( $L$ ) by equation (11).

3) Iteration with Gradient Descent method [10]:

$$V_0^{(k+1)} = V_0^{(k)} + \nabla W^{(k)}(u) * d(k) \tag{14}$$

Where  $V_0^{(k)}$  is the position of  $V_0$  after  $k$  times iteration.  $\nabla W^{(k)}(u)$  is a vector in the same direction as the gradient at  $V_0^{(k)}$ .  $d(k)$  is minimum search step. As  $\nabla W^{(k)}(u)$  is in parallel to the attachment of the last two control points, a unit vector in the same direction with  $\nabla W^{(k)}(u)$  can be expressed as:

$$\nabla W_{unit}^{(k)}(u) = \omega \frac{\overrightarrow{V_0^{(k)}V_1^{(k)}}}{\| \overrightarrow{V_0^{(k)}V_1^{(k)}} \|} \tag{15}$$

Where  $\overrightarrow{V_0^{(k)}V_1^{(k)}}$  is a vector from  $V_0^{(k)}$  to  $V_1^{(k)}$ .  $\omega = 1$  when  $L^{(k)} > L_0$  and  $\omega = -1$  when  $L^{(k)} < L_0$ . In order to improve the efficiency of calculation, set

$$\nabla W^{(k)}(u) = \nabla W_{unit}^{(k)}(u)(2 + \cos \theta_1) \tag{16}$$

Where  $\theta_1$  represents the angle of  $\nabla W^{(k)}(u)$  and  $\nabla W^{(k-1)}(u)$ .  $\theta_1$  can adjust the speed of calculating. When  $\theta_1 < 90^0$ ,  $\| \nabla W^{(k)}(u) \|$  will be bigger than 2, the calculating speeds up; When  $\theta_1 > 90^0$ ,  $\| \nabla W^{(k)}(u) \|$  will be smaller than 2, it slows down.

$$d(k) = \delta_2 L_0 \tag{17}$$

Where  $\delta_2$  is a positive constant much smaller than 1.

Repeat 2), 3) until meet the stop condition.

4) Stop condition for iteration:  $L - L_0 \leq \epsilon \delta_2 L_0$ .  $\epsilon$  is a constant greater than 1.

Figure 2 is a model gotten by this method. In the result of iteration, if the highly coordinate of  $V_0$  is less than 0, it is impossible that the free end is suspending, which implies that part of the cable contacts with the ground.

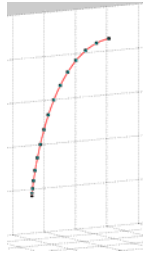


Fig. 2. A cable centerline model with free end suspending

### 4.2 Free End Dragging on the Ground

As friction with the ground is usually uncertain, deformation of the cable can not be exactly specified even if other constraints known. This paper proposes a new algorithm named Gradient Projection Descent method to search the point ( $V_{contact}$ ) where the suspending part contacting the ground, and take friction to the stop condition of calculation.

#### Influence of Relevant Factors

A cable with one end fixed and the other free on the ground, deformation of the suspending part is affected by friction ( $F_f$ ), gravity, bending stiffness parameter ( $\alpha$ ) etc.. Taking account of passing given points, set  $p = k \frac{F_f}{\alpha(G_L - \rho h^* g)}$ , where  $G_L$  represents the cable weight between  $V_{contact}$  and  $P^*$  ( $P^*$  is a given point nearest to  $V_{contact}$  in the suspending part,  $P^*$  is the fixed endpoint if there are no given points).  $h^*$  is the height of  $P^*$ .  $k$  is a constant for certain cable. Suppose maximum static friction coefficient here is  $f$ , then

$$p = k \frac{f(L_0 - L)}{\alpha(L^* - h^*)} \tag{18}$$

$L_0$  is the total length of the cable,  $L$  is length of the suspending part, and  $L^*$  is the length between  $V_{contact}$  and  $P^*$ . Then set a positive constant  $\gamma$ , and suppose that the deformation reaching a stable state when  $p < \gamma$ .

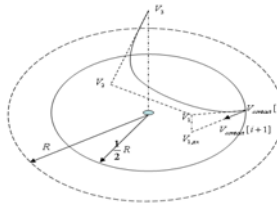
**Search with Gradient Projection Descent Method**

- 1) Initialize the starting point for searching.  
Get the position of  $V_0$  with the method of section 4.1 (Where the height of  $V_0$  is less than 0). Set the initial position of  $V_{contact}$  to the projection of  $V_0$  on the ground.
- 2) Take  $V_{contact}$  and  $V_3$  as the first and the fourth control point, search the position of  $V_1$  and  $V_2$ , and get the length of the cable ( $L$ ) (see section 4.1).
- 3) Iteration with Gradient Projection Descent method (as shown in figure 3):

$$V_{contact}^{(k+1)} = V_{contact}^{(k)} + \nabla W_{zx}^{(k)}(u) * d(k) \tag{19}$$

$\nabla W_{zx}^{(k)}(u)$  is a vector that consists of a same direction with the projection of  $\nabla W^{(k)}(u)$  (See section 4.1) on the ground. A unit vector that has a same direction as  $\nabla W_{zx}^{(k)}(u)$  can be expressed as

$$\nabla W_{zx,unit}^{(k)}(u) = \omega \frac{\overline{V_{contact}^{(k)} V_{1,zx}^{(k)}}}{\| V_{contact}^{(k)} V_{1,zx}^{(k)} \|} \tag{20}$$



**Fig. 3.** The sketch map of Gradient Projection Descent method

Where  $V_{1,zx}^{(k)}$  is the projection of  $V_1^{(k)}$  on the ground, and  $\overline{V_0^{(k)} V_{1,zx}^{(k)}}$  is a vector from  $V_0^{(k)}$  to  $V_{1,zx}^{(k)}$ . The definition of  $\omega$  is the same as section 4.1. Then

$$\nabla W_{zx}^{(k)}(u) = \nabla W_{zx,unit}^{(k)}(u)(2 + \cos \theta_2) \tag{21}$$

Where  $\theta_2$  is an angle between  $\nabla W_{zx}^{(k)}(u)$  and  $\nabla W_{zx}^{(k-1)}(u)$ .

$$d(k) = \delta_3 L^* \tag{22}$$

Where  $\delta_3$  is a positive constant that much smaller than 1.

When  $\nabla W_{zx}^{(k)}(u)$  is perpendicular to the ground, in order to avoid  $\nabla W_{zx}^{(k)}(u)$  set

$$V_{contact}^{(k+1)} = V_{contact}^{(k-1)} + 2 * \nabla W_{zx}^{(k-1)}(u) * d(k-1) \tag{23}$$

Repeat 2), 3) until meet stop conditions.

4) Iteration stop conditions:  $p < \gamma$  and  $L \leq L_0$ .

As the effect of the ground, the tangent of cable centerline at  $V_{contact}$  usually parallels the ground. Therefore this paper set the highly coordinate of  $V_1$  to 0 after obtaining the position of the four control points. Then rebuild the model of cable centerline with other coordinates unchanged. Connect the part on the ground with the part suspending at the point  $V_{contact}$  and keep the connection continuous in first derivative. See figure 4.

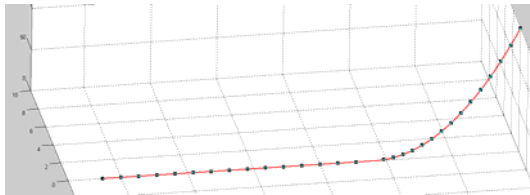


Fig. 4. A cable centerline model with part on the ground

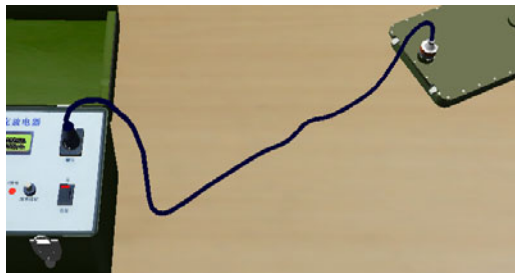


Fig. 5. A cable model in virtual maintenance training system

## 5 Conclusion

This paper divides the cables in virtual maintenance environment into two cases, one with both ends fixed, the other with one end fixed and the other end free. We adopt different algorithms for them, and get the centerline model for each one. As the algorithms consist of a small amount of calculation, real-time display can be achieved when constraints changes. So they can also be used to the process simulation of maintenance. We firstly tested these methods on MATLAB and got smooth cable centerlines. (See figure 1, 2 and 4.) Then we took it into practice in a virtual

maintenance training system and gained lifelike cable models (See figure 5). The studies of the paper enlarge the application of virtual modeling from rigid parts to large deformation parts, which constitutes a significant meaning for the development of virtual maintenance technology.

## References

1. Liu, P., Li, R., Hu, C.: Flexible Cable Modeling Method in Virtual Environment. *Journal of Ordnance Engineering College* 20(3) (June 2008)
2. Buck, M., Schomer, E.: Interactive rigid body manipulation with obstacle contacts. *Journal of Visualization and Computer Animation*, 243–257 (September 1998)
3. Sauer, J., Schomer, E.: A constraint-based approach to rigid body dynamics for virtual reality applications. In: *Proc. ACM Symposium on Virtual Reality Software and Technology*, pp. 153–161 (1998)
4. Ng, F.M., Ritchie, J.M., Simmons, J.E.L., Dewar, R.G.: Designing cable harness assemblies in virtual environments. *Journal of Materials Processing Technology* 107, 37–43 (2000)
5. Veltkamp, R.C., Wesselink, W.: Modeling 3D curves of minimal energy. *Proc. Computer Graphics Forum* 14(3), 97–110 (1995)
6. Moll, M., Kavraki, L.E.: Path Planning for Deformable Linear Objects. *IEEE Transactions on Robotics* 22(4) (August 2006)
7. Cui, T., Aiguo, S., Wu, J.: A Mass-Spring Model for Haptic Display of Flexible Object Global Deformation. In: *Proceedings of the 2007. IEEE International Conference on Mechatronics and Automation*, Harbin, China, August 5-8 (2007)
8. Huang, M., Dong, X.: Research on Flexible Cable Geometric Modeling Technology in Virtual Maintenance Based on VRML, 978-1-4244-7739-5/10 ©2010 IEEE
9. Celniker, G., Gossard, D.: Deformable curve and surface finite-elements for free-form shape design. *ACM Computer Graphics* 25(4), 257–266 (1991)
10. Chen, B.: *Theories and algorithms of optimization*, pp. 334–341. TsingHua Press

# Communication between Vehicle and Nomadic Device Used for Assistance and Automation

Daniel Thielen, Christian Harms, Tobias Lorenz, and Frank Köster

Institute of Transportation Systems, German Aerospace Center (DLR), 38108  
Braunschweig, Germany  
{daniel.thielen, christian.harms,  
tobias.lorenz, frank.koester}@dlr.de

**Abstract.** Introduction of wireless vehicular communications enables a variety of new Intelligent Transportation Systems (ITS) use-cases [1], [2], [3] allowing *cooperation* between vehicles and infrastructure services. Furthermore, modern vehicles feature an increasing number of driving *automation* functions. This paper identifies requirements of cooperative and automated ITS, with a special emphasis on *communication between a vehicle and a nomadic device*. An integrated ITS design is proposed by the German Aerospace Center (DLR) assembling *Vehicle Communications*, *Vehicle Technology* and *Infrastructure Technology*. This has been implemented in a prototype for evaluation in February 2011.

**Keywords:** Intelligent Transportation System (ITS) application, cooperation, automation, nomadic device, Vehicle Communication, Vehicle Technology, Infrastructure Technology, advanced driver assistance systems (ADAS), Vehicle-to-X (V2X).

## 1 Introduction

Development of future advanced driver assistance systems (ADAS) will most likely be driven by two key aspects: *cooperation* and *automation*. Cooperation means shared perception, decision making and action planning between several vehicles and traffic infrastructure (e.g. traffic light control) [1]. Automation refers to a high degree of automated control within a vehicle. Current advanced research projects even strive at fully automated driving, even in urban areas [4].

From a technical viewpoint, this trend is supported by the introduction of Vehicle-to-X (V2X) communication (IEEE 802.11p) and the utilization of emerging technological trends to realize new use cases (e.g. communication between vehicles and nomadic devices). The German Aerospace Center (DLR) currently builds an integrated testbed for ITS applications: Application-platform Intelligent Mobility AIM [5]. Among other things, this includes communication of mobile phones to public transportation as well as individual vehicles which may enable intermodal travel assistance.

One of the key challenges in the introduction of *cooperative* and *automated* systems will be resolving opposing requirements from those two parts. On the one hand, *flexibility* and *openness* are needed for cooperative systems to enable different business models. On the other hand, certain *safety* (e.g. functional safety [6]) level needs to be assured during the ongoing automation of ADAS.

For a better and more concrete understanding of conflicting requirements, a scenario which illustrates the integrated use of *cooperation* and *automation* in an ADAS is presented in Section “Problem Statement”. Specifically, this paper will then focus on sketching a solution in the field of *communication between a vehicle and a nomadic device* (see Section “Proposed Solution”).

## 2 Related Work

The development of solutions for vehicle to mobile device communication is advanced under different aspects in various projects. Daimler demonstrated challenges, integration aspects and future requirements for in-vehicle mobile device integration already on the MOST Interconnectivity Conference in Tokyo in 2005<sup>1</sup>.

Some of these aspects are also subjects in the Car Connectivity Consortium founded in march 2011. This consortium is further developing specifications and solutions for different use cases to improve the integration of mobile devices with in-vehicle infotainment<sup>2</sup>.

The projects MOBILIST [7] (focus on the urban area of Stuttgart) and DELFI [8] (german wide) focused the intermodal use of public transportation. As a result, different software products (e.g. EFA<sup>3</sup>) are available to provide a continuous intermodal route for a traveler. Furthermore the Service Oriented Architecture (SOA) Working Group researched in further development common service oriented architectures for usage in ITS with an emphasis on eSafety<sup>4</sup>.

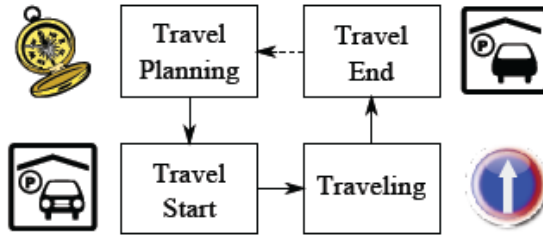
Additionally AIM [5] addresses the research interests of intermodal mobility.

## 3 Problem Statement

In this section, the aimed integrated scenario is sketched and the resulting requirements are derived and discussed.

### 3.1 Scenario

In order to show the possibilities of automation combined with V2I and nomadic devices communication, the main scenario is described as a sequence of four partial scenarios (see Figure 1): *Travel Planning*, *Travel Start*, *Traveling* and *Travel End*.



**Fig. 1.** Basic usage scenario

The given scenario was developed at the German Aerospace Center (DLR) and demonstrated to the public on February 2011 with the test vehicle *FASCar*.

### 3.1.1 Travel Planning

Travel planning consists of choosing the destination by the driver or traveler and calculating the route to navigate through, e.g. by a navigation service.

In the demonstration of February 2011, this task was computed offline. The map for the 2011 scenario including the route and four points of interest (POI) is shown in Figure 2.

### 3.1.2 Travel Start

When beginning the trip with a fully automated vehicle, the user just wants to leave the building and enter his car – without having to walk to a parking lot or garage. The car comes out of the parking lot automatically and waits in front of the building when traveling starts. In the demonstration of February 2011, at the beginning of the scenario, the passengers are located at POI 1 and the DLR’s research vehicle *FASCar* is located at POI 4 (see Figure 2).



**Fig. 2.** Demonstration scenario at the DLR research facility in Braunschweig (source: Google, <http://maps.google.de>, March 21st, 2011)

To start the scenario, one of the passengers calls the vehicle using the speech input of a nomadic device. As nomadic device a mobile smartphone has been used. Thereby a “Come Here” command is sent to the vehicle via UMTS communication. When receiving this command the vehicle reverses fully automated to the POI 1 where the passengers get on the vehicle.

### 3.1.3 Traveling

The actual travel is fully automated. The car follows the planned route, taking a detour if short-term changes of the traffic situation occur. It makes decisions about



maneuvers, and executes them. It reacts on the local traffic situation, e.g. following other cars, avoiding obstacles, changing lanes and stopping at traffic lights.

In the demonstration of February 2011, this part was implemented in the following way. First, the fully automated (longitudinal and lateral control) travel starts following the route and heading for the POI 2, where a cooperative traffic light is located. There, a *Traffic Light Assistance System (TLAS)* presents traffic light information to the passengers including a velocity advice, if reasonable. Reaching the POI 3, a second vehicle suddenly enters the route from the right, causing a hazardous situation. The research vehicle is able to detect the other vehicle and to stop automatically. Afterwards the travel is continued by a turning maneuver and heading back to the POI 2 where again the TLAS is presented.

### 3.1.4 Travel End

When the traveler gets to his destination, he does not want to spend time looking for a parking lot. He leaves the car at the destination and then it autonomously drives to a parking opportunity. In our scenario, when the *FASCar* reaches POI 4, the passengers leave the vehicle and the travel is finished. After that, it can be restarted again by the “Come Here” command.

## 3.2 Derived Requirements

From the usage scenario described in last section, requirements are derived in the categories *Human-Machine Interface (HMI)*, *Vehicle Communications*, *Vehicle Technology* and *Infrastructure Technology*. These requirements are used to design the ITS architecture and the functions.

### 3.2.1 HMI

Due to the strong connection between traveler and vehicle during the whole scenario, a HMI is needed in every partial scenario (see Subsection “Problem Statement-A”). Since in some of the partial scenarios the driver is outside the vehicle, a nomadic device is needed, which enables a communication with the automated vehicle. Whenever the traveler is in the vehicle furthermore the in-vehicle HMI is necessary.

### 3.2.2 Vehicle Technology

The automated vehicle shall be able to drive on roads, interact with other traffic participants, and obey traffic rules. It shall offer high-level services for valet parking, returning from parking, navigation and driving to the traveler. For this, reliable and accurate positioning and environment perception is needed alongside driving maneuver planning and vehicle control.

### 3.2.3 Communication Technologies

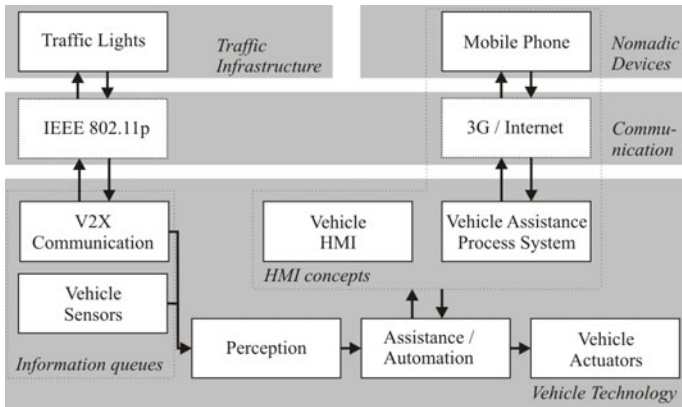
For the vehicle to support the sketched scenario, communication between the vehicle, the traffic infrastructure (e.g. traffic lights) and the traveler’s nomadic device shall be possible. It has to be assured that the communication with the nomadic device is not limited to a certain distance around the vehicle, so that it is always possible for the traveler to communicate with the vehicle.

## 4 Proposed Solution

In the following a solution is proposed to successfully fulfill the described scenario. First, a system architecture with minimal set of necessary components is designed. After that, further subsections give details on the different parts focused on *Communication* and *HMI concepts (Nomadic Device)*.

### 4.1 System Architecture

In Figure 3, the overall system architecture is presented. Based on the requirements from the previous subsection, four main architecture elements were identified: *Traffic Infrastructure*, *Nomadic Devices*, *Communication* and *Vehicle Technology*. These main architecture elements were detailed with concrete modules as a basis for the realization of the aimed functionality.



**Fig. 3.** Architecture of the overall system

As part of the traffic infrastructure currently the traffic lights are considered. These traffic lights are equipped with a communication device to exchange data with the vehicle using the IEEE 802.11p standard. The details on the traffic light technology are not focused by this paper but can be found in other publications. Because of emerging mobile technology trends extending to automotive research (e.g. foundation of the Car Connectivity Consortium<sup>5</sup>) a mobile phone is used to connect the traveler outside the vehicle. For communication between the mobile phone and the vehicle a service oriented architecture (SOA) is applied. A Vehicle Assistance Process System (VAPS) prepares and transmits it thru a 3G-Internet connection to a Vehicle Management & Information Server (VMIS) located on a Webserver where the data is accessible by a Webservice.

Regarding the vehicle, a sequence starting with information queues and ending up in the vehicle actuators and HMI concepts is presented. The information queues represent the sensors within the vehicle as well as the communication queues, gathering environmental information to be used by the perception block. Within the perception block the collection, interpretation, processing and assessment of the

gathered data for the use in the assistance and automation systems is performed. Subsequently in the assistance and automation block control signals are calculated which are used by the vehicle actuators to control or to assist the control of the vehicle in longitudinal and lateral driving direction. To build up a communication from the assistance / automation with the driver, the vehicle HMI is employed.

This paper focuses on *Communication* and *HMI concepts* as elementary parts for enabling a data exchange between mobile devices and vehicles.

For the described scenario it is necessary to develop a distributed solution: The first step deals with a travel planning. The standard action to plan a trip deals with assembling different information from different sources like timetables of public transportation (train, metro, trams, busses) and flights as well as traffic and real time traffic information. Resulting from different projects in the past, much of this information is already used on Internet pages or in apps for mobile devices.

Without referring to the details of such solutions one important aspect is the lack of individual transport data (e.g. car data). A reason for this is a missing interface to request individual car data. In this paper an approach for solving this problem is presented: a collaborative system, based on service oriented architectures with its parts *Vehicle Assistance Process System (VAPS)*, *Vehicle Management & Information Server (VMIS)* and *Nomadic Device*. This solution enables the integration of individual car-data in different use cases like the communication between a vehicle and a nomadic device used for assistance and automation.

### 4.2 Communication

The *Communication* element comprises the technologies that are used to exchange data between the vehicle and external hardware like nomadic devices or cooperative traffic infrastructure. To assure that the communication is not limited to a certain distance around the vehicle or nomadic device, this solution was chosen. To establish a connection, a 3G/UMTS device is applied which enables an internet connection over a mobile communication provider. This approach requires a sufficient cell reception which was available within the demonstrated scenario.

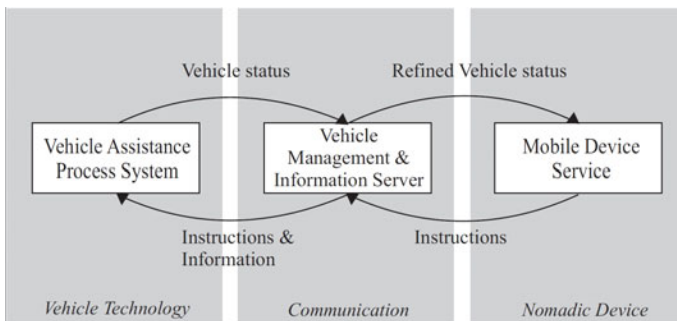


Fig. 4. Distributed systems exchange data between vehicle and nomadic device

#### 4.2.1 Process

The process of data-exchange between vehicle and nomadic device follows the representation in Figure 4. The Vehicle Assistance Process System provides FASCar status data (e.g. position information) by using a Webservice running on the Vehicle Management & Information Server (VMIS). The VMIS stores incoming data on a database and provides them by another Webservice. Due to the described scenario a mobile device consumes this service. It transmits control signals (in this solution a flag for starting fully automated driving) back to the VMIS. This data is maintained on the VMIS as long as another query from the Vehicle Assistance Process System takes place and it will be returned as part of the confirmation.

#### 4.2.2 Collaborative Systems

Communication between the vehicle and a nomadic device is implemented as a service oriented architecture (SOA) with its main parts the *Vehicle Management & Information Server (VMIS)* located on a Webserver and the *Vehicle Assistance Process System (VAPS)* located on the FASCar (see Figure 4). Vehicle Assistance Processes transmit vehicle information like the actual GPS position via Webservices to the VMIS. Instructions, stored on the VMIS by the users nomadic device, are fetched the same way by the vehicle.

#### 4.2.3 Vehicle Assistance Process System (VAPS)

The Vehicle Assistance Process System consists of an Application Server with an integrated BPEL (Business Process Execution Language) Container and a portfolio of different services giving access to vehicle data and functions. BPEL is appropriate to describe a business process consisting of different Webservices usually available on the Internet as well as in organizations. In this case, the service is provided on a vehicle. The BPEL container periodically transmits FASCar data to the VMIS by using a Webservice hosted on it and receives a control-flag status as confirmation. This status comprises the information about starting fully automated driving. If this applies the control-flag-status will transmit to the vehicles Assistance / Automation (see Figure 3).

#### 4.2.4 Vehicle Management and Information Server (VMIS)

To provide the data published by the Vehicle Assistance Process System a Vehicle Management & Information Server was implemented which also includes an Application Server. Two Webservices based on SOAP and WSDL are available on this system. The Simple Object Access Protocol (SOAP) is a XML-based protocol to exchange data over an Internet protocol (e.g. HTTP). The Web Services Description Language (WSDL) provides the model for the detailed service description.

On the VMIS one service receives the vehicle status information from the VAPS and stores it in a database. After that a confirmation is sent with an extension of the current control-flag-status. The control-flag-status is made available by another Webservice which receives control signals from a client (in this solution the control-flag-status is transmitted by the nomadic device) and provides FASCar status data (e.g. position information). Consequently this control-flag-status will be processed by the Vehicle Assistance Process System and hand out to the vehicle in order to start fully autonomous driving.

### 4.3 HMI Concepts – Nomadic Device

The overall architecture consists of four main elements. The *HMI concepts* comprises the HMI in the FASCar and on the nomadic device as well as a module for data-transfer between them. This subsection will focus the process and the implementation of the nomadic device application.

#### 4.3.1 Android Platform

In regard to a fast software development for mobile devices the Android platform was selected to realize the travel planning software. Additionally a rich set of tools (e.g. OS, middleware, and several development tools) are available including an extensive documentation to nearly all featured properties of the operating system and a great variety of supported hardware.

Android is an operation system and software platform for mobile devices and tablets. It is open source and is developed by the Open Handset Alliance under the control of Google. Because of its open source keynote, Android applications will be developed in Java. Furthermore they are constructed to follow the component-based software engineering and users can share their applications among each other [9].

#### 4.3.2 Application Flow Chart

The general functionality of the implemented Android application is presented in the flow chart in Figure 6. The software is used on a HTC Desire HD smartphone. In the main activity (simplified for screen in an Android-application) it is possible to start the described scenario via speech input. Therefore it is necessary to open the application's submenu via pressing the Menu-Button. The submenu opens and the user can choose the options *speech command*. As a result, the application starts a speech request. Following successful completion (a command has spoken) the application starts the speech recognition. This functionality was realized with the Android library for speech recognition which is available since Android Version 2.2 library is available. The implemented recognition reacts on the agreed command "come here". If this order is detected by the software, the application consumes the Webservice and transmits the flag for starting automated driving. After that, the user receives a confirmation of this step on his mobile device. Consequently this information is saved on the Vehicle Management & Information-Server and demanded by the described Vehicle Assistance Process System running on the vehicle.

#### 4.3.3 Implementation

At the Institute of Transportation Systems at DLR an intermodal travel assistant for mobile devices was developed on this platform. In combination with general information about travel start (comprises location and time) and preferred public transportation also individual car data (data from DLR's research car FASCar) are applied in this application to provide a continuous itinerary for the user. After the computed route is displayed, the user can decide to get information about the travelling route or receive status information from the FASCar. Due to the requirements of the described scenario, the developed travel application was simplified to its basic functionality: the exchange of data between nomadic device and

the DLR’s research car. To achieve this, a service oriented architectures was developed (IV-B) and a Webservice bases on SOAP provided. To consume the Webservice in the android application it is necessary to use external libraries or to implement an own Client and Parser because Android does not support SOAP based web services directly. To avoid dependencies from third-party applications the described solution is based on the second opportunity. By default, the Android SDK comes with different opportunities to open content from the web. Android provides an HTTP client interface which encapsulates objects required to execute HTTP requests. In short, this solution is used together with a self-implemented parser to interact with the Webservice and to handle the results. The implemented application bases on the architecture in Figure 5.

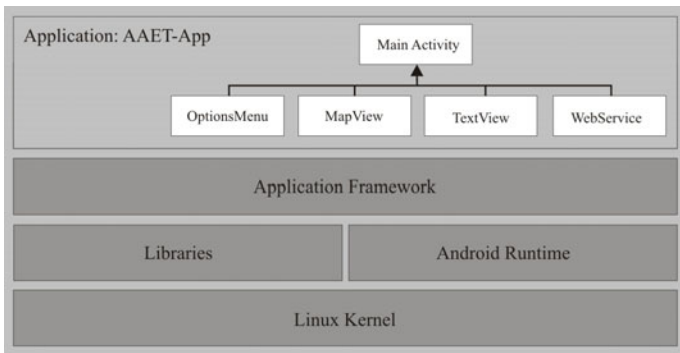


Fig. 5. Android Application - Architecture for the described scenario

Common Android applications as well as the implemented software base on a *Linux Kernel* (here: Version 2.6) and use the *Linux* components security, process management, network and driver model. The *Linux Kernel* serves as abstraction level between hardware and software components. Android includes different *libraries* (e.g. a media library for playing or recording Mediafiles). The *Android Runtime* contains the Dalvik virtual machine to execute each application in a separate instance of it. Additionally the Runtime includes a set of core libraries which are similar to usual core libraries of the Java programming language. Coming along there are specific Android libraries for different use cases, for example the screen painting. Developers can access these libraries via the *Application Framework*.

The actual application contains one main activity [5]. An activity combines a graphical user interface (GUI) with application logic. The activity uses an optionsmenu (i. a. for activation the speech input), a class called *Webservice* which handles the interaction between the Vehicle Management & Information Server and provides the data to show the current position of DLR’s research car on a Google map (MapView). Additionally, a *TextView* is used for general information.

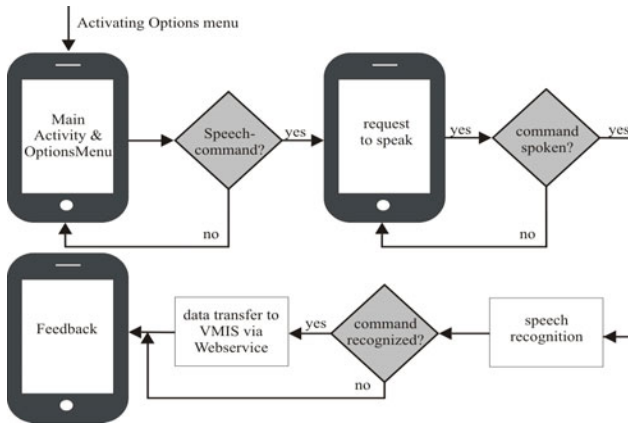


Fig. 6. Android Application Flow chart

## 5 Evaluation and Conclusion

In this paper the implementation of a system to enable a data exchange between nomadic devices and vehicles is presented. After the description of the needed hardware components the requirements of such a system are outlined and a possible solution is presented. When evaluating the system within a demonstration scenario III-A the overall functionality of it could be proved. In summary, the main conclusions of this paper are:

- Service oriented architectures (SOA) are an appropriate solution for realizing an interoperable system in an automotive context
- The Vehicle Assistance Process System (VAPS) and the Vehicle Management & Information Server confirm their suitability in enabling data-transfer between a vehicle and a nomadic device State of the art mobile smartphones based on Android are a potential platform to develop innovative ITS solutions
- The integration of business tools (BPEL) in a vehicle environment which supports new possibilities of this technology was successful
- The introduced solution provides scope for additional research in ITS use cases with an emphasis on interoperability between vehicles and mobile phones

To validate the presented concepts and the technical stability as well as the user acceptance, it is necessary to evaluate the system. Therefore the next is to bring a demonstration scenario similar to the one described here to a regular urban traffic environment. This will be enabled through an integrated testbed for ITS applications, which is currently built by the German Aerospace Center (DLR): AIM [5] (Application Platform Integrated Mobility). Among other things this testbed includes an equipment of the urban traffic infrastructure with V2I communication devices.

**Acknowledgments.** The authors would like to thank the government of Lower Saxony as well as the Helmholtz foundation for funding the depicted research at the German Aerospace Center (DLR).

## References

1. ETSI TR 102 638 – Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions, V1.1.1 ed., European Telecommunications Standards Institute (ETSI) (June 2009)
2. simTD Deliverable D11.1 – Beschreibung der C2XFunktionen, V1.0 ed., Sicher Intelligente Mobilität – Testfeld Deutschland (June 2009)
3. SOA Working Group, Final report of the Service Oriented Architecture Working Group (2010)
4. Wille, J.M., Saust, F., Maurer, M.: Stadtpilot: Driving autonomously on Braunschweig's inner ring road. In: 2010 IEEE Intelligent Vehicles Symposium (IV), pp. 506–511 (2010)
5. Köster, F., Hannibal, M., Frankiewicz, T., Lemmer, K.: Anwendungsplattform Intelligente Mobilität (AIM) - Dienstespektrum und Architektur. In: AAET 2011. Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel (2011)
6. ISO/DIS 26262: Road vehicles – Functional safety, International Organization for Standardization (ISO) (2009)
7. Günther, C., et al.: Mobilität im Ballungsraum Stuttgart - Das Leitprojekt MOBILIST - Ergebnisse, Evaluation, Umsetzung (Auszug aus dem Abschlußbericht des MOBILIST-Konsortiums) (2003)
8. Schnitger, S., et al.: DELFI - deutschlandweite elektronische Fahrplaninformation, Report on the current status of the DELFI-system implementation (2006)
9. Google Inc., What is Android? (2010), <http://developer.android.com/guide/basics/what-isandroid.html> (visited on December 20, 2010)



# Texture Synthesis Based on Wang Tiles

Yan Liu<sup>1,\*</sup> and GuangLin Cheng<sup>2</sup>

<sup>1</sup> Qingdao Hotel Management College, Qingdao, Shandong, China  
<sup>2</sup> Huawei Technologies Co., Ltd, Nanjing, Jiangsu, China  
Qingdao\_teacher@163.com, Orange200610@163.com

**Abstract.** Texture Synthesis aims to solve the existing problem of seams in texture mapping. It is used to generate an arbitrarily large texture image given an existing small texture sample. For human observers, both the original texture image and the generated one are perceived as the same and the generated one does not contain obvious seams. Non-periodicity is an import principle in texture synthesis. Wang Tiles algorithm is known to synthesize texture according to a color matching rule via designing a Wang Tiles set. We believe if we utilize more tiles and the number of edges of each tile increases, that means, we use more colors, we can reduce periodicity in the synthesized results. The chosen texture tiles in the previous algorithms are mainly rectangular. Combined with Wang Tiles and Quilting algorithms, in this paper we design a hexagonal texture tile. Through the experiment our method is effective and not time-consuming.

**Keywords:** Texture synthesis, Wang Tiles, hexagonal tile.

## 1 Introduction

Texture is mostly described as the Markov Random Fields[1]. Efros and Freeman adopted the Quilting method to achieve the 2D texture synthesis based on texture tiles[2]. They chose smaller rectangle tiles from the sample textures randomly and sewed them up in the overlapping areas, which greatly improved the synthetic speed and is better than the synthesis based on pixel according to its result. Texture synthesis algorithm which is based on tiles has become the current focus of research. Wang Tile's synthesis algorithm[3][5][7] adopted the method of designing square texture, via color matching of the tiles' sides to achieve texture synthesis. And the tile's number of design must assure at least two optional tiles of each synthesis position in order to achieve the non-periodicity of the texture synthesis. Kwatra and other researchers achieved texture synthesis of the image and video, using GraghCut's method. During the process, they did not end the experiment with only once synthesis as before[2][3][5][7], otherwise, they constantly found the bigger error places and optimized the seams problem[4]; Meanwhile, they proposed the conception of closed path to generate a new texture. Literature [6] focused on data structure. What is input did not only include one sample image, but a piece of noise image. The tree structure

---

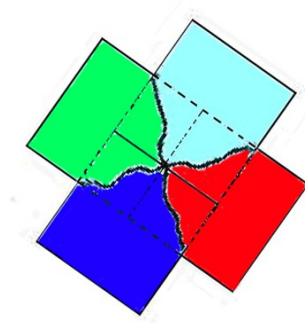
\* Corresponding author.

was adopted in the texture synthesis and obtained a good result. Besides, it could make the synthetic speed raise two orders of magnitude than the former algorithm. Different from the previous methods which needs only one sample texture, Lefebvre, Han and other researchers [8][9] generate high-resolution and arbitrarily large texture images based on several low-resolution image by using a Gaussian Stack . Through the process of sampling, jittering, rectification, they achieve the multi-scale texture synthesis, which can not only form infinite, large-scale texture image, but also can observe slight details of the result texture. This can't be achieved in the past. Li-Yiwei and other researcher thought oppositely and proposed the theory that[10]: if the objective theoretic texture is given, how we chose the smallest tile which included all the information of the given texture and generated the objective theoretic tile. Of course, researchers have done a lot of researches on texture synthesis, such as the newest solid texture synthesis[11]. The former texture syntheses are either 2D texture synthesis in the plane or 3D surface texture synthesis, but the solid texture synthesis can go into the inner part of 3D tile and acquire a good result. Different from the methods in [3][5][7],if the designed textures have more sides and colors, it will certainly lower the periodicity of the synthesized result, which is the motive of this paper's research.

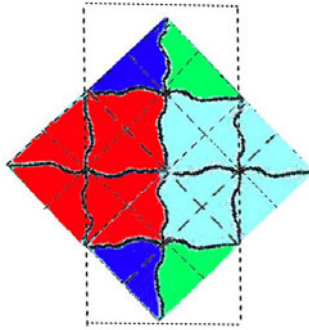
The rest of this paper is organized as following. Section 2 mainly focuses on introducing the algorithm of designing the hexagonal texture elaborately. Section 3 introduces the texture synthesis algorithm using the color matching rule. Section 4 makes a comparison of the experimental results. Finally, the conclusion of this whole paper is given in Section 5.

## 2 How to Design the Hexagonal Tile

We combine the methods of [3][5] to design our needed hexagonal tile. Firstly, we choose four tiles with  $45^\circ$  angle according to the established threshold. Then we try to find out the best cutting path on the diagonal direction using the method in [5], which can generate a rectangle whose sides are different in color. As shown in figure 1, according to color matching rule of contiguous sides, we combine the four generated tiles together and there are no seams within the tile. We cut the hexagonal part in the middle of it as the tile we want. We can see from figure 1 that through the designing process as described above, the six sides of the hexagonal tile used in the experiment can have different colors in order to eliminate the periodic of the final synthesized result. To be mentioned, the rectangle tiles at the horizontal sides of the hexagonal must have the same color, which is to prove that that same synthesis is used when the rectangle is designed in order to assure that the designed hexagonal sides have the same color, otherwise, the joint of different colors might cross the vertical line of the cutting and the synthesized texture tile will have obvious seams. What's more, both horizontal sides are different from other sides and they are diagonals. So when we design the hexagonal texture tile, we must pay attention that the left and right sides must be synthesized using the same texture tile. That's why the inner part of the two tiles in figure 2 has the same color. As shown in figure 2, the hexagonal in the dashed box is the one we get via the method mentioned above. During the texture synthesis process, we choose twelve colors. That means we have twelve texture tiles.



**Fig. 1.** How we design a rectangle tile with four colors by rotating  $45^\circ$



**Fig. 2.** How we obtain our hexagonal tile by combining four tiles. The tile in the dash rectangle is what we want.

### 3 Texture Synthesis Using Hexagonal Tiles Based on Wang Tiles

#### 3.1 Texture Synthesis Using Hexagonal Tiles

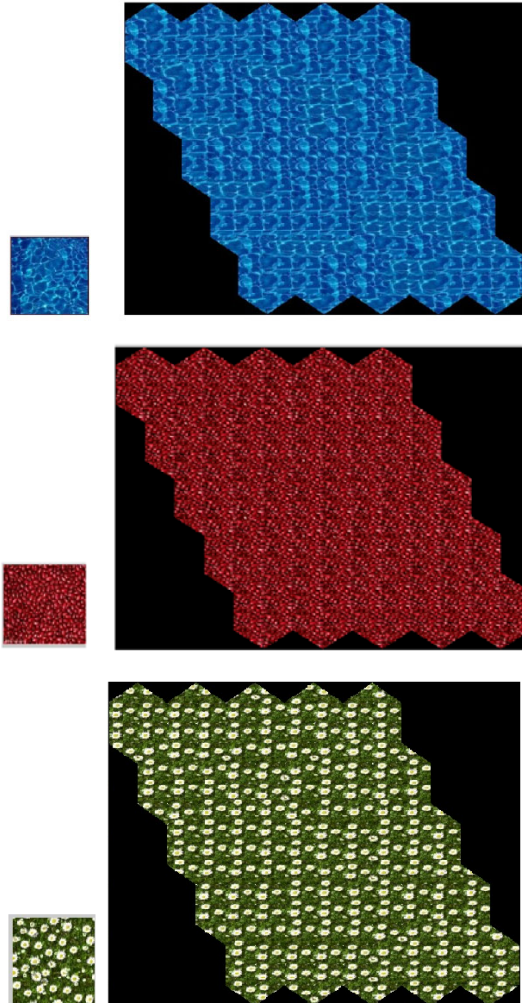
Different from Wang Tiles algorithm [3][5], we do not design a tiles set, but produce the hexagonal tile while synthesizing the result texture. That is, we firstly match the colors, forcing that the adjacent edges have the same color. Therefore we eliminate obvious seams. The left colors of edges are selected randomly from all twelve colors. At the very start, we set a data structure for storing the color values of each hexagonal tile and they are all initialized zeros. For each step we must store the color values of current hexagonal tile.

1. For the first row we randomly select six colors to design the up-left corner tile;
2. For other tiles of the first row, we only need to force that the left color value of new tile equals to the right color value of previous tile, and we randomly select the other five color values;
3. If it is not the first row, for the first tile we need match the upper color values. Otherwise, we should consider the left color value together;

We continuously execute the second and third steps until we obtain the result texture of required size.

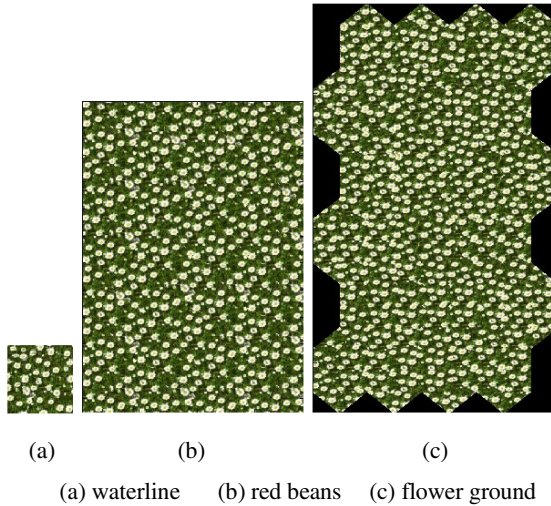
### 3.2 Experimental Results

As shown in figure 3, we can easily get the synthesized results randomly via using hexagonal tiles of arbitrary size. Because of the restriction of resolution, we just give the synthesized results of  $5 \times 5$  tiles.



**Fig. 3.** The synthesized result using hexagonal tiles  
Left: sample texture; Right: synthesized texture

As in figure 4, compared with Wang Tiles algorithm, the synthesized results we obtain have a lower periodicity meanwhile we can finely eliminate the seam problem, for we adopt more rectangle tiles so that we have more selectivity of colors.



**Fig. 4.** Comparison between Wang Tiles algorithm and ours

## 4 Conclusion and Implication

Combined with the Quilting algorithm, this paper brings forward a new algorithm, which is based on Wang-Tiles algorithm, to design hexagonal tiles. Then we utilize the designed tile to synthesize texture and acquire good results. However, this algorithm also has some limitations to improve. The authors wish to design a hexagonal tiles set in future to improve the calculating efficiency of this algorithm.

**Acknowledgments.** The authors give thanks to Junyu Dong for his help.

## References

1. <http://baike.baidu.com/view/405080.htm>
2. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: International Conf. on Computer Vision, pp. 1033–1038. ACM Press, Greece (1999)
3. Cohen, M.F., Shade, J., Hiller, S., et al.: Wang tiles for image and texture generation. *ACM Transactions on Graphics* 22(3), 287–294 (2003)
4. Kwatra, V., Schodl, A., Essa, I., et al.: Graphcut Textures: Image And Video Synthesis Using Graph Cuts. *ACM Transactions on Graphics*, 277–286 (2003)
5. Burke, R., <http://www.heroicsalmonleap.net/mle/wang/> (August 2003)
6. Wei, L.Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: *Proc of ACM SIGGRAPH PH*, pp. 479–488. ACM Press, Los Angeles (2000)

7. Wen, C., Tan, T.-S., Zhang, X., Kim, Y.J.: Generating an  $\omega$ -tile set for texture synthesis. *Computer Graphics International*, 177–184 (2005)
8. Lefebvre, S., Hoppe, H.: Parallel controllable texture synthesis. In: *SIGGRAPH*, pp. 777–786 (2005)
9. Han, C., Risser, E., Ramamoorthi, R., Grinspun, E.: Multi-scale texture synthesis. *ACM Transactions on Graphics* 27(3) (2008)
10. Wei, L.-Y., Han, J., Zhou, K., Bao, H., Guo, B., Shum, H.-Y.: Inverse Texture Synthesis. In: *SIGGRAPH* (2008)
11. Takayama, K., Okabe, M., Ijiri, T., Igarashi, T.: Lapped solid textures: filling a model with anisotropic textures. In: *Proceedings of ACM SIGGRAPH*, vol. 27(3) (2008)

# Research on the Control Strategy of the Driving System for a Kind of Hybrid Electric Vehicle

Lu Fang<sup>1,2</sup>, Wang Deng-feng<sup>1</sup>, and Li Da-peng<sup>2</sup>

<sup>1</sup> State Key Laboratory of Automotive Simulation and Control, Jilin University,  
Changchun 130022

<sup>2</sup> Technical Development Department, FAW-Volkswagen Automotive Company, Ltd.,  
Changchun 130011

**Abstract.** The features of driving system, the parameter matching and the control strategies of hybrid electric vehicle have determined the dynamic performance, the fuel economy, the emission characteristics, the manufacturing cost and the weight of the vehicle. In this paper, a kind of hybrid electric vehicle is considered as the research object and the control strategy of the driving system of this hybrid electric vehicle has been analyzed. Eventually, based on the parallel auxiliary driving, the system parameters have been selected according to the relevant system standards and the requirements for performance. At the same time, on account of the SOC torque balance control strategy, the simulation model of the whole system has been established by Matlab/Simulink. This simulation has verified the feasibility of the design and the effectiveness of the control strategies, which has achieved the desired goals. Compared with the traditional vehicles, this hybrid electric vehicle has improved the fuel economy and reduced the emissions.

**Keywords:** Hybrid electric vehicle, control system, control strategy, system simulation.

## 1 Introduction

With the continuing decline of oil resources and the increasing serious environmental pollution, the energy conservation has become the primary task for the all walks of life. The car manufactures and the research institutions have paid more and more attention to the green-car with low emission and low fuel consumption. China has clearly put forward the strategy of new energy vehicles in “The Twelfth Five-Year Guideline”. As a kind of new energy vehicles, the hybrid electric vehicle is not very different from the traditional vehicles and it is easy to be implemented. At the same time, it also has some advantages such as the long work hours, the good power performance, the less pollution and the low noise, which is the important development direction for the fuel-efficient vehicles. The design objective of the hybrid electric vehicles is to achieve the rational design of driving system and realize the parameter matching for the components of driving system as well as the coordination control of vehicles in the process of driving. On account of a kind of hybrid electric vehicle, the parallel auxiliary driving system has been established in this paper. Meanwhile, the





The power transfer route of this driving system can be shown in Figure 2. The arrows in this Figure refer to the directions of power transfer. P1 is the output power of engine. P2 refers to the charging power for the batteries when the engine load is small. P3 is the auxiliary power provided by the assist system. P4 refers to the output power of torque coupled device to the driving system.

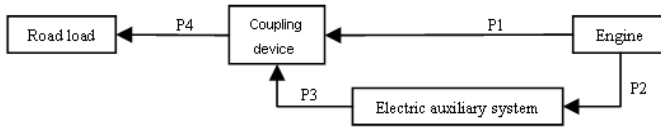


Fig. 2. Power transfer route of system

### 3 The Control of Driving System

#### The Control Principle of Parallel Auxiliary Driving System

The parallel auxiliary driving system can receive the control signals and the vehicle status information through the vehicle controller. And the control strategies can be employed to control and coordinate the whole driving system. This system requires that the vehicle controller can reasonably distribute the torque among the engine, the motor and the brake according to the current power needs and the driving conditions. Meanwhile, the controller should be also able to select the starting time of engine and motor. When the demanded torque of system is negative, it is necessary to control the motor so as to recycle the regenerative braking energy as much as possible. When the demanded torque is positive, it is required to control the torque output ratio of the two power sources.

#### The Control Strategies of Parallel Auxiliary Driving System

For the parallel auxiliary driving system, the development of control strategies is a global multi-objective optimized process, which should take into account of the fuel economy, the engine emissions, the battery life, the dynamic performance and the cost requirements, etc. At the same time, on account of the system characteristics and the operating conditions, the engine, the motor, the battery and the transmission system should be achieved the best matching [5]. At present, the control strategies of the parallel hybrid system mainly include Electric Assist Control Strategy, SOC Torque Balance Control Strategy, SACS Self Adaptive Control Strategy and FLCs Fuzzy Logic Control Strategy [1].

Energy management and torque regulation are mainly concerned in this paper, so the SOS torque balance control strategy is selected.

### 4 The Setting of the Control Parameters of Driving System

#### The Performance Index of Vehicles

On account of the development of domestic auto market and the hybrid electric vehicles, the performance of the hybrid electric vehicles studied in this paper should achieve the following aspects: continuous mileage should be reached to 300km; the vehicle can run 30km only driven by the engines (the average speed is 50km/h); the

maximum speed is 140km/h; the acceleration time is less than 16s; the largest climbing gradient is 30%.

**The Parameters of Parallel Auxiliary Driving System**

The hybrid electric vehicle in this paper is the front-driving sedan which has four doors and two covers. The parameters of this vehicle shown in Table 1.

**Table 1.** Cutting parameters of simulation milling

Dimension parameter	Length (mm)	Width (mm)	Height (mm)	Axle spread (mm)
	4650	1760	1700	2650
Mass parameter	Kerb mass (Kg)	Mass except transmission system (Kg)	Load mass (Kg)	Rate of axle load
	1380	890	300	0.54
Characteristic parameter	Coefficient Of drag	Coefficient of rolling resistance	Airflow area (m <sup>2</sup> )	Radius of tire
	0.3	0.12	2.2	0.305

The calculation of power in accelerated driving can be shown in formula 1:

$$P_{max} = \delta M_v V_m \left[ V_m - V_m \left( \frac{t}{t_m} \right)^{0.58} \right] / 0.1 + 0.5 \rho_a A C_d V_m^3 + M_v g V_m C_r \tag{1}$$

In this formula,  $M_v$  is the vehicle quality;  $V$  is the speed;  $\rho_a$  is the air density;  $A$  is the frontal area;  $\delta$  is the quality equivalent coefficient;  $C_d$  is the drag coefficient;  $C_r$  is the rolling resistance coefficient.

The calculation of power in uniform climbing can be shown in formula 2:

$$P_g = 0.5 \rho_a A C_d V_g^3 + M_v g V_g \sin \theta + M_v g V_g C_r \cos \theta \tag{2}$$

In this formula,  $\theta$  is the climbing angle.

The calculation of battery capacity  $C_b$  can be shown in formula 3:

$$C_b = \frac{P_e}{\eta_{tra} \cdot \eta_{mot}} \cdot \frac{S_e}{V_e} \cdot \frac{1}{SOC_{high} - SOC_{low}} \tag{3}$$

In this formula,  $S_e$  is the driving distance in pure electric mode.  $V_e$  is the average speed.

Considering the losses of the components of driving system in power conversion and transmission, the efficiency  $\eta$  has been introduced to calculate and round  $P_g = 42kW$ ,  $P_{max} = 64kW$  and  $C_b = 6.36kW.h$ . Eventually, the four-stroke gasoline engine has been selected. The maximum output power of engine is no less than 42kW and the maximum output torque is no less than 85Nm. It is required to

select the AC induction motor. The maximum power is 25kw, the maximum speed is 9000rpm and the rated speed is 2250rpm. At the same time, it is necessary to choose the NIMH battery pack and the capacity should be between 6kWh and 8kWh.

### 5 The Modeling and the Simulation of Control System

#### The Model of Control Strategies

In this paper, the Matlab/Simulink is used to setup the model for each system module of vehicles. The diagram of the simulation model is shown in Figure 3.

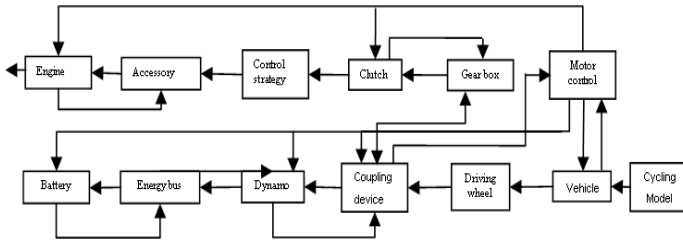


Fig. 3. SOC torque balance control strategy model

According to the previous description, this hybrid electric vehicle has selected the SOC torque balance control strategy. And the simulation model of control strategy can be shown in Figure 4.

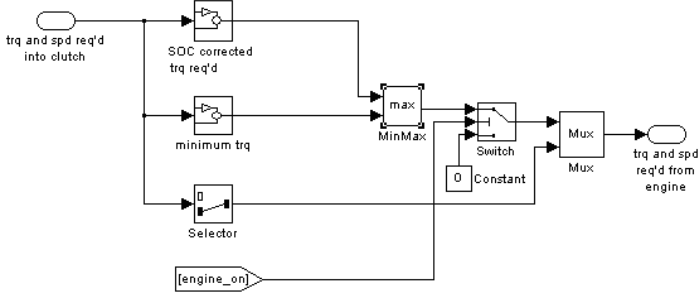
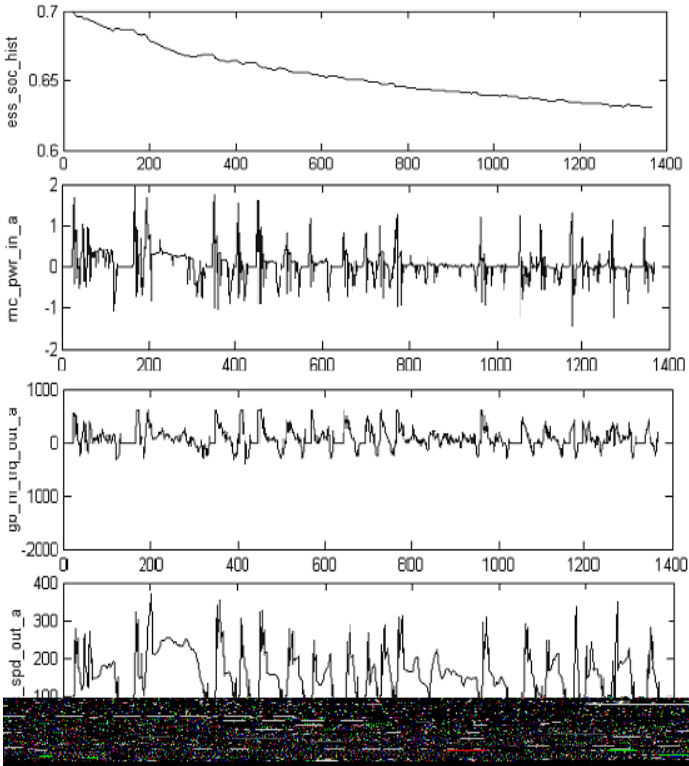


Fig. 4. SOC torque balance control strategy model

The SOC torque balance control strategy refers that the engine has been controlled to send the torque according to the SOC condition and the torque ratio. In this paper, we have set that the SOC correction coefficients can be automatically corrected so as to get the torque value of engines in different situations.

#### The Performance Simulation of Vehicles

According to the parameters, the UDDS road condition is considered as the input to carry out the simulation analysis on the fuel economy, the emissions and the vehicle dynamic performance of this hybrid electric vehicle under the influence of SOC torque balance control strategy. The simulation results can be shown in Figure 5.



**Fig. 5.** Results of SOC torque balance control strategy

Simulation results prove that the cycle will be successfully completed by employing this strategy. The engine speed and the changes of SOC value are both within the permitted scope. The simulation results have achieved the desired objectives.

**The Comparison of Simulation Results**

The comparison of the fuel economy, the emission data and the dynamic performance between the hybrid electric vehicle and the traditional vehicle under the influence of SOC torque balance control strategy can be shown in Table 2.

**Table 2.** Target compare with traditional vehicle

Control strategy	Climbing ability (slope/80km/h)	Accelerated index (s)	0~100	0	0~100 km/h
Hybrid electric vehicle		17.8%	11.4	7.4	24.1

From these data it can be found that the hybrid electric vehicle has achieved performance target of the traditional vehicles. The employment of SOC torque balance control strategy has met the performance requirements and improved the fuel economy as well as reduced the pollution emissions. Therefore, it is successful for this hybrid electric vehicle to adopt this strategy.

## 6 Conclusion

On account of a kind of hybrid electric vehicle, the feasibility that the parallel auxiliary driving system is appropriate for this vehicle has been analyzed in this paper. The components have been selected and the system parameters have been designed. According to the relevant system standards in our country and the influence of battery on the dead weight of system, the dynamic performance parameters of this vehicle have been determined and the reasonable matching of driving system has been also realized.

Through the study, it is confirmed that there is a certain reliability of the SOC torque balance control strategy for this hybrid electric vehicle. At the same time, the corresponding simulation model has been established in Matlab/Simulink and the method which combines with the forward simulation and the backward simulation has been employed. Through the simulation results, the effectiveness of the control strategies of this driving system and the feasibility of the design program have been verified, which has achieved the pre-set dynamic performance objectives and saved time for the following development works. Compared with the traditional vehicles, the hybrid electric vehicle has improved the fuel economy and reduced the pollution emissions.

## References

1. Johnson, V.H., Wipke, K.B., Rausen, D.J.: HEV Control Strategy
2. Federal Motor Vehicle Safety Standards. Title 49 of the United States Code, Part 571
3. GB T18386-2001. The Requirements of Electric Vehicles
4. Chen, C.C.: The State of the Art of Electric and Hybrid Vehicles. Proc. of the IEEE, 90(2) (2002)
5. Marshaus, J., Ramnarine, A., Brodsky, E., et al.: Development of the University of Wisconsin's Parallel Hybrid Electric Sport Utility Vehicle. SAE FutureTruck Challenge, 167-181 (2001)
6. Gonder, J., Markel, T.: Management Strategies for Plug-In Hybrid Electric Vehicles. SAE Paper 2007-010290
7. Chatzakis, J., Kalaitzakis, K., Voulgaris, N.C., et al.: Designing a New Generalized Battery Management System. IEEE Transactions on Industrial Electronics 50(5), 990-999 (2003)

# The Wireless Ranging System Based CSS Technology

Sun Hao, Wang Xin, Lv Changzhi, and Gong Weimin

College of Information and Electrical Engineering, Shandong University of Science and  
Technology, Qingdao, China

sun5782@163.com, xinxinjiayou@126.com,  
lvchangzhi@126.com, gong5783@163.com

**Abstract.** This paper designs a wireless measuring system by the application of CSS technology and the use of NanoPAN5375 RF transceiver module and AVR mega1280 microcontroller, which achieves the long-distance, high-precision design requirements. Because the both ends can read data after the communication, this method can reduce much field work. By measuring the system has realized the technical program requirements, which has the features of low power, portable, anti-interference ability, and has a high value and market prospects in the actual production.

**Keywords:** CSS technology, wireless ranging system, NanoPAN5375, AVR mega1280.

## 1 Introduction

In recent years, the wireless sensor network technology has made rapid development, including the information collection and transmission. In the practical application process, the network nodes are not static, so the node positioning is very important. How to measure inter-node distance is the basis of positioning and ranging accuracy will directly affect the positioning accuracy. Currently, GPS technology is most widely used in the location technology, but because of the cost and application environment constraints, it is not very realistic to configure the GPS for every node of wireless sensor networks. So it is very necessary to set up an accurate distance measuring system for wireless sensor networks.

This paper designs a wireless ranging system by the application of CSS technology and the use of NanoPAN5375 RF transceiver module and AVR mega1280 microcontroller. Dual ranging bilateral symmetric algorithm SDS-TWR using the CSS technology, can achieve an accuracy of 1m The wireless ranging system based on the CSS technology can achieves the long-distance, high-precision design requirements, and has the features of low power, portable, anti-interference ability, and has a high value and market prospects in the actual production.

## 2 Nomenclature

### 2.1 Wireless Ranging Principle

Measuring distance by electromagnetic waves has mainly two ways: one is RSSI technology based on the received signal strength used to determine the strength of the distance between two nodes. Because of the different media encountered in signal propagation which results in different attenuation, the accuracy of RSSI technique will be not very high. The other is the use of signal propagation time between two nodes to calculate the distance. So obtaining the signal propagation time is the key to this method, and it has relatively high ranging accuracy.

The first method is TOF. We can directly get the distance airborne time of the ranging signals. This ranging schematic is shown in Figure 1.

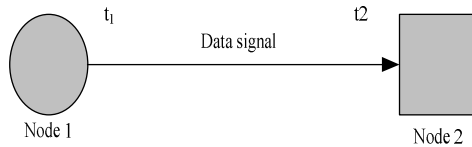


Fig. 1. TOF ranging schematic

Based on propagation speed  $v$  known in the medium, and the time measured by the formula (1), we can get the distance  $d$  between two nodes.

$$d = v \cdot (t_2 - t_1) \tag{1}$$

But this method requires the same accurate clock of two nodes which strictly keep pace, so it is difficult to implement, and the cost will increase. In order to overcome the clock difficulties, it presents dual ranging bilateral symmetric algorithm SDS-TWR<sup>[1]</sup>. This method can effectively overcome multi-path transmission and noise interference, and improve range accuracy. This ranging schematic is shown in Figure 2.

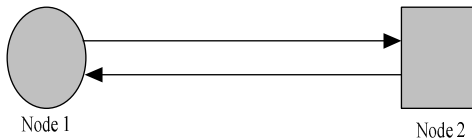
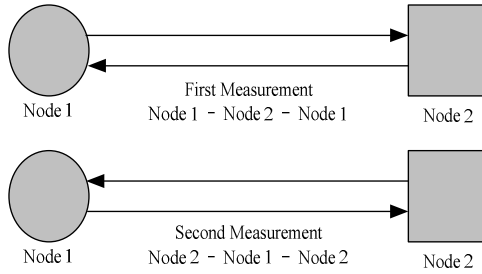


Fig. 2. TWR ranging schematic

TWR ranging process is the process. First, Node1 sends ranging signal to Node2. Then, Node2 returns a signal to Node1.

In order to overcome the impact of clock drift and improve the range accuracy, it makes two symmetrical measurements which is bilateral symmetry measurement SDS. In the normal measurement mode, NanoLOC chips use the SDS-TWR algorithm.



**Fig. 3.** SDS-TWR ranging schematic

The implementation process is shown in Figure 3. In the first measurement Node1 records the time  $T_1$ , and gets the recording time Node2 of  $T_2$  by ANSWER1; In the second measurement Node1 records the time  $T_4$ , and gets the recording time Node2 of  $T_3$  by ANSWER2. The difference between  $T_1$  and  $T_2$  is the first time between the nodes. The difference between  $T_3$  and  $T_4$  is the second time between the nodes.

$$t = \frac{(T_1 - T_2) + (T_3 - T_4)}{4} \quad (2)$$

In the SDS-TWR algorithm, the signal has four transmissions between the nodes. According to the formula (2), it can calculate the one-way signal transmission between the nodes. Then according to the signal propagation velocity in the air, it can get the distance between nodes.

## 2.2 Hardware Design

Hardware circuit consists of three parts: AVR mega1280 microcontroller for system control module, NanoPAN5375 wireless transceiver module and the external circuit.

### (1) Microcontroller and the External Circuitry

The ranging system requires a microcontroller to control the entire system, and the microcontroller's performance will determine the performance of the system, so it is very important to choose a suitable micro-controller. To achieve this distance, the system requires micro-controller which has a little speed and processing power, and has enough data storage space for data computing and temporary. Besides, as a mobile device, it requires a lower power consumption micro-controller.

The performance of mega AVR microcontroller series is not only superior, but also a very good cost performance. The system uses AVR mega1280 as the system micro-controller.



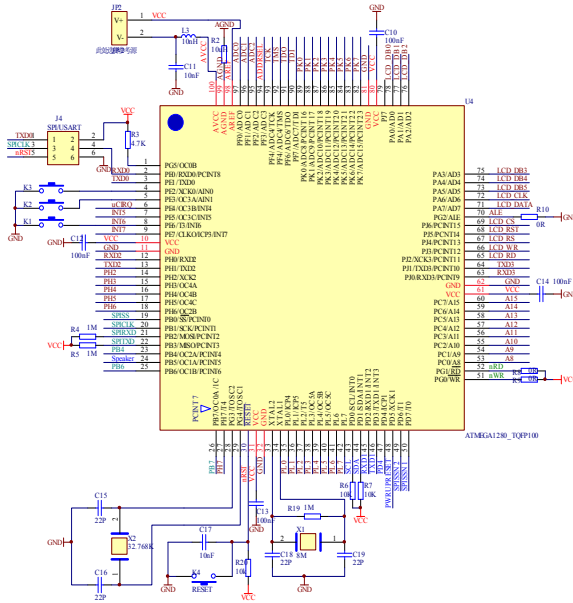


Fig. 4. Control module

(2) NanoPAN5375 RF Module and the External Circuit

NanoPAN5375 module is a band based 2.4GHzISM and the RF module of integrated amplification, filtering and other components. It integrates all necessary components, so it streamlines its peripheral circuits. In addition, it also needs to add an antenna. NanoPAN5375 RF module and its external circuit are shown in Figure 5.

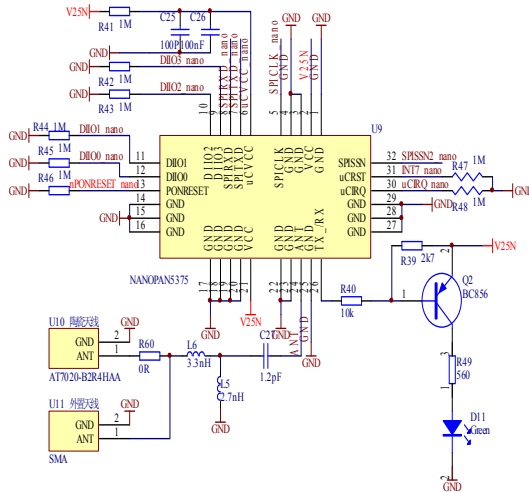


Fig. 5. NanoPAN5375 RF module and its external circuit

NanoPAN5375 RF module uses the SPI interface to communicate with the microcontroller. SPI is a synchronous serial peripheral interface, which allows a variety of SPI serial peripheral devices to communicate to exchange information by the master-slave manner. SPI has four signal lines, respectively, the serial clock SCK, master input slave output signal MISO, master out slave input MOSI and slave select signal SS. AVR microcontroller also has a SPI interface, which easily makes single-chip radio module to communicate with NanoPAN5375.

### (3) Port Connection between Microcontroller and RF Module

The power supply voltage of single-chip system is 3.3V, while NanoPAN5375 RF module power supply voltage is 2.5V, so the microcontroller and RF modules NanoPAN5375 ports don't directly connected. The main microcontroller communicates with NanoPAN5375 RF module via SPI interface, but SPI bus signal timing has very strict requirements. So the voltage conversion circuit of the microcontroller and RF modules NanoPAN5375 can not have too much delay, or timing can not be normal traffic. 74HC245 is a high rate tri-state buffer bus driver, which designs for two-way synchronous communication data bus. It is a CMOS chip and compatible with TTL level, and can achieve high-speed voltage conversion. 74HC244 control consuming is very small, and signal delay is at the ns level. So this method can achieve voltage conversion between the microcontroller and RF modules NanoPAN5375.

## 2.3 Software Design

NanoPAN5375 supports these sets of point-to-point wireless communication protocol control, address matching, prior to the check, modulation mode, data transfer rate and other parameters set. The messages of the PHY hardware adaptation layer Software driver code, directly send to NanoLOC chip by the controller's SPI bus, and provide automatic address matching and irregular packet retransmissions<sup>[2]</sup>. NanoLOCnTRX is the main function chip within NanoPAN5375 RF module, provides the structure and method calls of software driver package.

The paper designs a lost child-resistant system. Program flow chart system is shown in Figure 6. First, it completes measurement to get the distant of two nodes; Second, according to the data, it makes the processing applications to determine whether the distance is safe or not. So the system begins to perform different operations.

In the process of debugging the system, it is not stable at a value obtained from the data in several measurements to the same system, but there are obvious fluctuations, which directly affect the system measurement distance accuracy, and increase the system's false alarm. Therefore, it is necessary to make the filtering after obtaining the distance data in the system.

The design uses a simple average filter algorithm: it designs a queue to store 20 data. When it measures every distance, the distance will press new information into the queue, and the oldest data stored in the queue will be deleted, so this queue is always the most recent 20 times ranging results. The data is obtained as the current distance by averaging these 20 data. After debugging, the performance has been improved greatly.

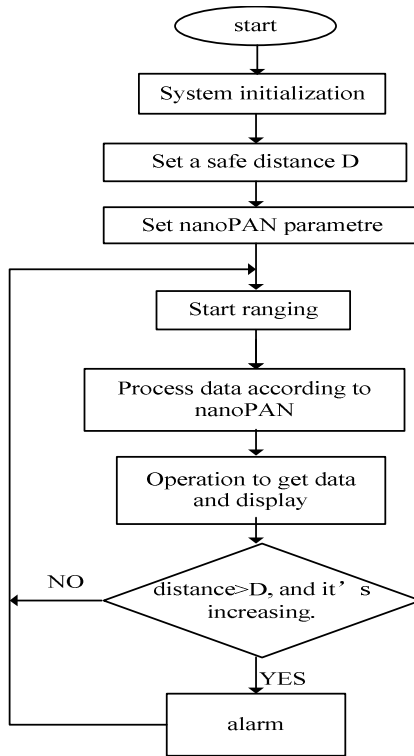


Fig. 6. Program flow chart

### 3 Test Results

Because of the limits of the standard measurement tools, it makes the actual measurement with 20 sampling points in 150 meters. Figure 8 shows the measuring data is in the standard range, and it basically meets the requirement. The error will be stable at about 5%. When the distance is larger, the error will be smaller.

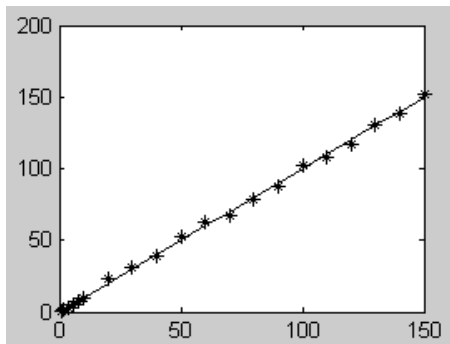


Fig. 7. Measurement proofreading figure

Dual ranging bilateral symmetric algorithm SDS-TWR using the CSS technology needs battery power and requires high data rate sensor networks<sup>[3]</sup>. CSS technology has the very strong anti-interference ability with good anti-effect on the path, and even uses the radio frequency echo. The CSS high system gain, and the advantage of anti-noise, interference and attenuation, can achieve a longer transmission distance in the same power output and operating conditions. Using the CSS technology, carrier frequency and data clock can be synchronized in an instant, and quickly establish wireless connections. CSS system has the ability of measuring the distance between two nodes, then it can turn location-aware.

The ranging system applying the AVRmega1280 microcontroller and wireless transmitter modules NanoPAN5375 has the features of rapid data transmission, low power, portable features. Because the both ends can read data after the communication, this method can reduce much field work. And has a good practical value and market prospects in the industrial control, civil project, many aspects of home life.

## References

1. Nanotron Technologies GmbH, Real Time Location Systems (RTLS) (2006)
2. Nanotron Technologies GmbH, nanoLOC TRX Transceiver (NA5TR1) Datasheet-Version 2.3 (2009-09)
3. Nanotron Technologies GmbH, nanoNET Chirp Based Wireless Networks-Version 1.04 (2007-01)

# Research on Communication Network of Smart Electricity Utilization for Residence Community

Zhang Xin-lun<sup>1</sup>, Wang Zhong-jian<sup>1</sup>, Bu Xian-de<sup>2</sup>, and Zhu Yu-jia<sup>3</sup>

<sup>1</sup> Jiangsu NARI Yinlong Cable CO. LTD., Beiyuan Road, China, Xuzhou, 221700

<sup>2</sup> State Grid Electric Power Research Institute, No. 8 NARI Road, China, Nanjing, 210003

<sup>3</sup> Hohai University Power and Electrical Engineering Institute, No. 88 Huju Road, China, Nanjing, 211100

**Abstract.** Optical Fiber Composite Low-voltage Cable can used to transfer energy, television, telephone and data(for internet) information, by this means, the communication network infrastructure could be built once a time and be shared by different service providers, so State Grid Company has regulated a series of standards on how to construct communication network for smart electricity utilization Residence community. The paper introduced the theories and key technology firstly, and then it illustrated a kind of solution on network and analyzed network security issues. At last it discussed the necessary policy support to further application.

**Keywords:** Smart Electricity Utilization, Communication Network, Optical Fiber Composite Low-voltage Cable, EPON Technology, Three-Network Convergence.

## 1 Introduction

As the development of science and technology, the advanced communication and information control techniques have not only changed our life in different ways ,but also altered people's life ideas and living habits, therefore living conditions are improved greatly. Meanwhile, people's living conceptions have changed that the requirements for living are not limited in area, surroundings and traffic, they began to seek smart, in formalized, humanized residence with clear air and clean energy. Under this position, Smart electricity utilization residence community came into being.

The smart electricity utilization residence community generally used smart ammeters and intelligent interactive terminals to build connections with users and to realize info collection and devices control. It can lower the labor cost, electricity waste and transmission loss, enhance metering precision, ensure power supply, promote energy saving and reduce ejection. The services that should be supported involved electricity utilization information gathering, utilization and distribution automation, FTTH, multi-networks convergence, renewable energy, photovoltaic power generation, smart house etc.,.

## 2 Construction Proceedings

American announced to build smart utilization community in 2008 at Boulder Distinct, each family has setup a smart ammeter to control electricity consumption in

a reasonable way, at the same time, it can access the clean energy such as wind power and solar power with priority, the residuary energy can be transferred to substations nearby, so other families can get this power from the power grid. Xcel Energy Power Company has planned to try to use smart control techniques and business modes to operate about smart grid.

Holland started working the smart city construction in 2008. It aimed to realize zero growth of CO<sub>2</sub> ejection and began with intelligent electricity utilization, LED street lamps reconstruction, intelligent buildings and ammeters, electromotor charging station, solar electricity generation and such 17 tentative projects.

Japan has decided to make new communities with renewable energy supply. Kyushu and Okinawa Power Company has made many experiments to use solar power energy in local area, and Japan Economy Department planned to spread the use this kind of “Island tiny grid” in a large range in future.

In North Korea, the intelligent utilization techniques are the main investment on developing green energy in further 20 years, which can increase ratio of green technology form 2.4% to 11%. Kepco hopes to help residences to manage electricity by making use of power line broad band communication techniques and smart ammeters via different gateways. Besides that, it also intended to found a sample project about smart power grid before 2011, and the power utilization grid should be accomplished in 2020 and applied in the whole country in 2030.

In China, State Grid Company has expanded the research on smart electricity utilization community and business operation modes form 2005. In 2010, some research institutes were organized to working out the standards on communication networks of smart communities. These standards have regulated how to establish and mange the devices and systems relatively to convergence different kinds of communication networks and utilize renewable energy in the whole country.

### **3 Communication Network Framework Used in Smart Communities**

Communication network should provide the transmission channels for the information of smart grid, electromotor controllable data, public communication services, which can be shared with different applications as network infrastructures. Besides broad band communication services that can be offered for users, the investment could be reduced and advanced services such as “Internet of Things” be available in the future.

The purpose of the network is to accomplish FTTH (Fiber To The Home: to provide fiber channel access with extreme broad band), which is to say that using Optical Fiber Composite Low-voltage Cable(OPLC) as physical transmission medium in low-voltage communication access network to connect with ammeters or terminals. The fiber could be composed and paved along with power cable to load the gathered electric power grid information, electricity utilization information and telecom services by the use of EPON (Ethernet Passive Optical Network) technology.

The communication network for smart electricity utilization community logically could be two parts: electric power private sub-network and pubic communication sub-network. The private network is mainly to transmit electric distribution automation system, electric utilization information gathering system, smart energy consumption system, distributed power supply management and electric power application systems.

The other one can used to apply public services, such as 95588 website, television, telephone or internet services.

## 4 Basic Principles and Key Technologies

### 4.1 Basic Principles

After analyzing the demand that smart utilization services required, the communication network framework logically can be divided three layers according with application, which are device access layer, communication device layer and service layer. Figure 1 shows the framework of the FTTH communication network.

Device access layer provides different ways to connect all kinds of terminal devices of FTTH network. These devices could be smart ammeters, intelligent electric appliances, sensors, High-definition television (HDTV), IP phone (also be named with VoIP), internet device, water meter or gas meter etc.

Communication Network device layer is consisted of power fiber network devices to apply EPON technology and support two directions data flow in one single fiber. These devices are of the three kind: OLT (Optical Line Terminal: To provide interfaces between IP network and local exchange and to communicate with ONU devices), ONU (Optical Network Unit: To connect with OLT and to provide Internet surfing, VoIP, HDTV, Videoconferences to users) and ODN(Optical Distribution Node: to provide fiber transmission channels for OLT and ONU devices).

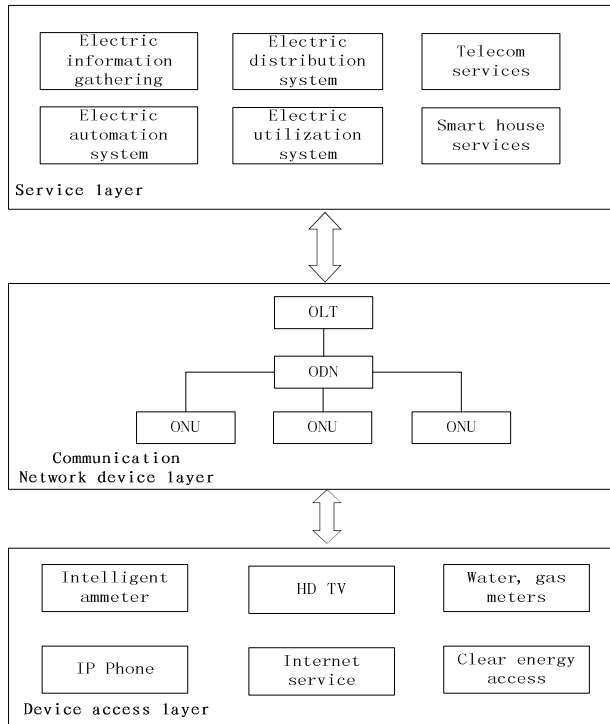


Fig. 1. Logic framework of FTTH communication network

Service layer is basically to bear operation system for smart electricity utilization community.

## 4.2 Optical Fiber Composite Low-Voltage Cable Technology

OPLC relatively productions[2] are primarily designed to transmit electricity with metal cable and to carry optical signal with fiber; the cable and fiber are compounded into one line together. By this means, the power grid and communication network can be finished synchronously. OPLC products are principally applied in 0.6/1kV or lower voltage grade distribution and utilization power grid.

As we can see, it could reform power and telecommunication network resources, share the infrastructure, avoid repetitive construction of different networks (electric line, internet wire, telephone and CATV). Figure 2 shows the structure of OPLC cable.

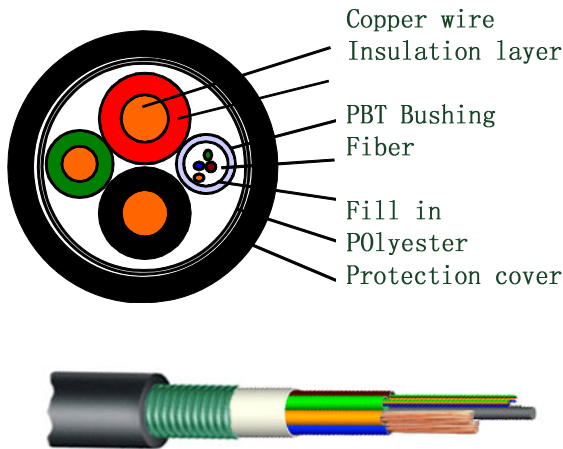


Fig. 2. Structure of OPLC

The cost of manufacturing this OPLC cable will increase 10%~20%, compared with traditional cables, but the total cost to build power and telecommunication network using OPLC cables will decreased by 40% approximately.

## 4.3 EPON Technology

EPON (Ethernet Passive Optical Network) is a new kind of optical access network. It adopts point to multipoint topology and passive optical fiber transmission technology, can provide multiple services upon Ethernet. EPON uses passive optical network technology on physical layer and Ethernet protocol on data link layer. EPON technology integrates the advantages of PON and Ethernet. So it has the characteristics of low cost, high broadband and strong expansibility[3].

Being a major access technology of FTTH, the EPON technology is developed on the basis of the Ethernet technology. It enables seamless integration with IP and



Ethernet technologies. Compared with other optical network access technologies, the EPON technology obtains more popularity due to its fine scalability, simplicity and multicast convenience.

A typical EPON system consists of OLT, ONU and ODN. The OLT is located at the central office and is used to control channel connection, management and maintenance. The ONU is used to access end users or floor switches. Through a single optical fiber, data from multiple ONUs can be multiplexed to an OLT port via ODN. The point-to-multi-point tree topology structure is adopted, which reduces investments on convergence equipment and results in clearer network hierarchy [4].

### 5 Solutions

The FTTH communication network can be used for power private services and public communication services. Each service needs a separated EPON system and the two kinds of services should be isolated by firewalls-like sets which should ensure the security of data transmission[5].

Figure 3 gives a typical solution about how to build communication network for smart electricity utilization community.

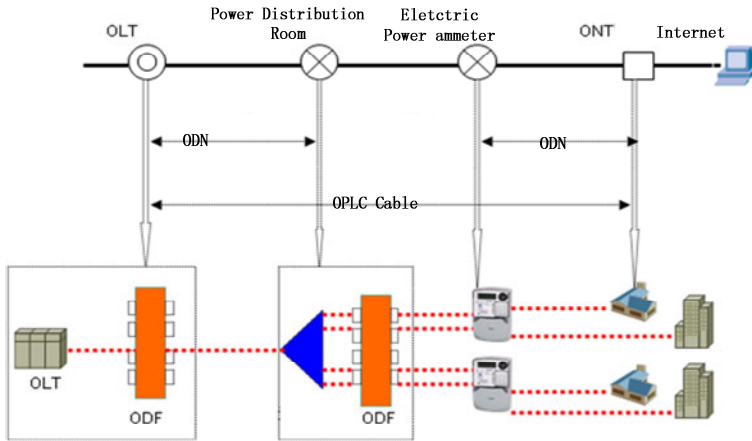


Fig. 3. A solution about network project

#### Network Security

This kind of network has such security troubles as following:

- 1) Unauthorized illegal access;
- 2) Data intercept and capture;
- 3) Malicious code attack, Denial of service;
- 4) Device configuration illegal modification;

The network security policy should include network environment, computer system and applications to protect the data transmission and avoid to be destroyed or attacked by hackers or virus. In order to intensify the security and integrality of the whole system, authentication is necessary for the users or terminals to ensure the uniqueness and authenticity.

## 6 Issues on Popularization and Application

The community construction requires the government to make policies or standards to build and manage the network, but the most important thing is to coordinate electric power enterprises with government, users, service providers and other partners relatively to get win-win situation, maintain the system operating, lower the cost and highlight efficiency.

**Acknowledgments.** This research project is sponsored by China State Grid Electric Research Institute.

ZHANG Xin-lun, Male. Bored in 1963. Professor of SGEPRI. zhangxinlun@sgepri.sgcc.com.cn.

WANG Zhong-jian, Male, Professor of SGEPRI. wangzhongjian@sgepri.sgcc.com.cn

BU Xian-de, Male, Engineer, Research on telecommunication for electric power system. buxiande@sgepri.sgcc.com.cn.

Zhu Yujia, female, (1990-) studying in Hohai University Power and Electrical Engineering Institute. She assists in the research of electrical engineering and automation.

## References

1. Deng, G.-P., Chen, J.: Smart Electric Utilization Community and Key Technology. Hubei Power S1, 21–24 (2010)
2. Li, X., Yang, J.-H., Qi, X.-W.: Analyze and Discuss about OPLC Application. Telecommunications for Electric Power System, 24–27 (July 2011)
3. Zhang, H., Bu, X.-D., Guo, J.-H.: Application of EPON technology used in Electricity Utilization Information Gathering System Telecommunications for Electric Power System 07, 42–45 (2011)
4. Liu, N.-G., Liu, D., Yang, S.-Z.: Application of EPON Technology Used in Smart Electricity Utilization Community. Hua Dong Power 03, 32–36 (2011)
5. Ding, J., Sun, J.-C.: Study on Communication Platform Used in Smart Electricity Utilization Community. Power Supply and Utilization 06, 8–12 (2010)

# Security Audit Based on Dynamic ACCESS CONTROL

Li Weiwei, Zhang Tao, Lin Weimin, Deng Song, Shi Jian, and Wang Chen

Institute of Information and Communication, State Grid Electric Power Research Institute,  
Nanjing, 210003, China  
b03040604@126.com,  
{zhangtao, linweimin, dengsong,  
shijian2, wangchen2}@sgepri.sgcc.com.cn

**Abstract.** The paper proposed a design of security audit based on dynamic access control, which periodically generated rules by mining latest log and applied rules to the packet filtering. Based on matching the content of packet with rules to measure network behavior security, this design generated different mechanism of security access control. Comparing to existing systems which made constant rules to respond to mutative filtering, this design had great applicability and flexibility.

**Keywords:** data mining, security audit, access control.

## 1 Introduction

The rapid development of network technology brings not only convenience, but also threats to network security. Network security system requires that the consequence of network behavior is predictable and controllable.

The main content of the security audit is to analysis log and use it for prevention, investigation and analysis. Most of the existing security audit system research used constant rules to respond to mutative network flow by the way of date mining, which made the system lack of flexibility and applicability [1-2].

The paper presents a security audit mechanism of dynamic access control, which can dynamically generate rules based on mutative system network access by measuring the network behavior security during the packet filtering. This mechanism improves the flexibility of access control, realizes access control and achieves the purpose of preserving network.

## 2 Dynamic ACCESS CONTROL

### 2.1 Formal Description

The process of realizing dynamic access control relates to generation and conversion of multiple tables.

For further description, a few definitions will be introduced first.

The definition of rule table: Rule table is a rule set which generates by data mining and expressed as  $R = \{R_1, R_2, \dots, R_n\}$ . Every item  $R_n$  describes the characteristics of a rule. Rule table is the basis and foundation to achieve dynamic access control.

The definition of cache table: Cache table is a set of data flow recently and expressed as  $C = \{C_1, C_2, \dots, C_n\}$ . Every item  $C_n$  describes the assessed value and characteristics of data flow. Assessed value is the measure reflection of matching data flow with rule tables.

The definition of white list: White list table is a set of date flow which can be trusted without matching with rule table and expressed as  $T = \{T_1, T_2, \dots, T_n\}$ . Every item  $T_n$  describes the timestamp and characteristics of data flow. The timestamp is the basis of eliminating expired list.

The definition of black list: Black list table is a set of date flow which can be denied without matching with rule table and expressed as  $B = \{B_1, B_2, \dots, B_n\}$ . Every item  $B_n$  describes the timestamp and characteristics of data flow.

## 2.2 The Process of Dynamic Access Control

The dynamic access control adjusts assessed value dynamically and measures the list for access control of date flow recently.

The dynamic access control implements through the following three aspects.

(1) The rule table is generated and issued periodically. The log is extracted and analyzed to generate new rules by data mining based on the truth that the data flow is the same in the adjacent time intervals. Rules are issued by the process and used in packet filtering. The purpose is to prevent the old rules form controlling new data flow.

(2) Black lists and white lists are generated dynamically. Black lists and white lists are generated by measuring the assessed value in catch table by matching packet content with rule table lists. When the assessed value reaches white assessed value, the list will be added to white list. The list can be trusted and pass directly for a period. It is the same to black list table. It can improve the filter efficiency.

(3) Black lists and white lists are removed periodically. The update process scans the black list table and white list table and removes the expired lists according to the timestamp. The purpose is to respond to mutative filtering.

The process of forming dynamic access control is shown in figure 1.

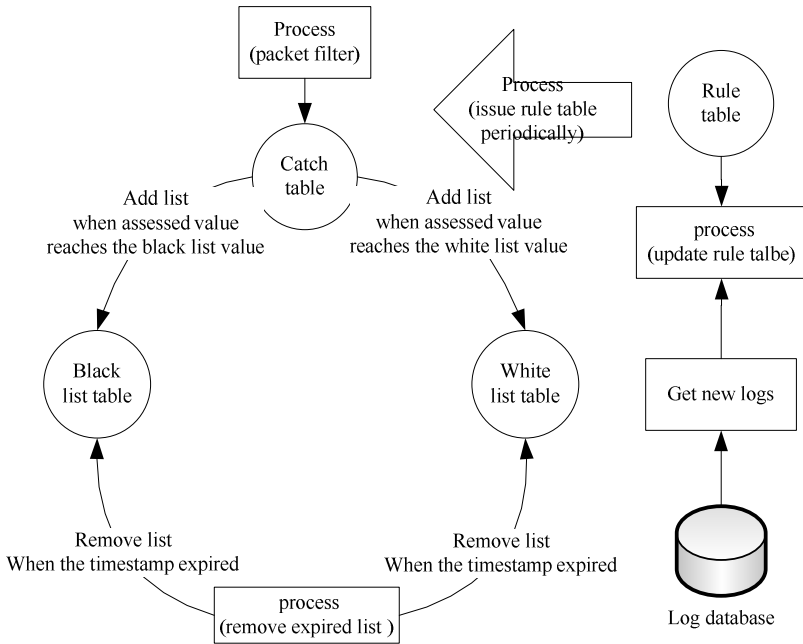


Fig. 1. The process of dynamic access control

### 3 Security Audit Based on Dynamic ACCESS CONTROL

#### 3.1 Architecture

This paper proposed a design of security audit system based on dynamic access control.

The main architecture of the security audit system model based on dynamic access control is shown in figure 2.

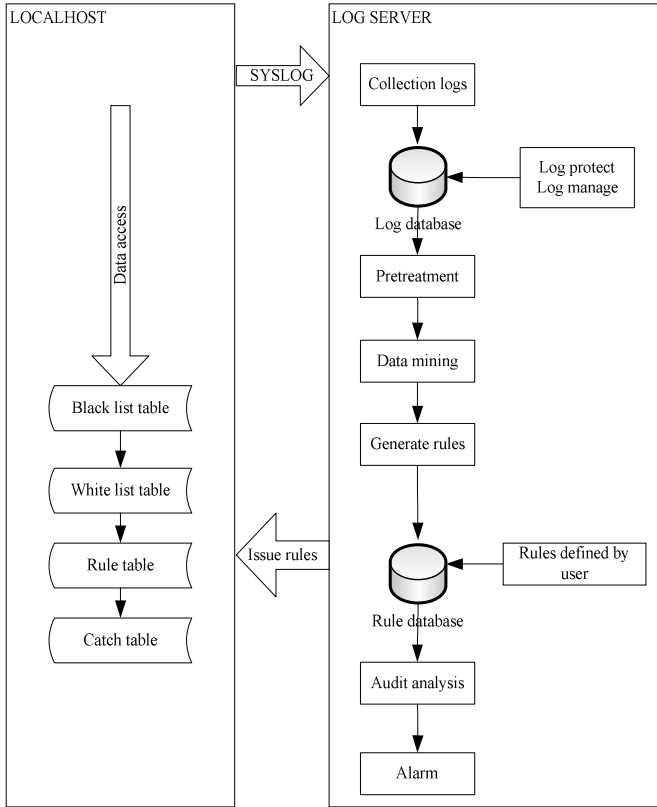


Fig. 2. The security audit system model based on dynamic access control

The main flow of the security audit system is described as follows:

Step (1): The system builds safe tunnel between local host and audit service.

Step (2): After the tunnel has been build, the log is send by the mechanism of syslog form local host to log server.

Step (3): The server analyzes the recent log and generates rules by data mining periodically.

Step (4): When the new rules has been generated, the process in server issues rules to local host.

Step (5): In local host, the packet is matched with different tables and assessed for the packet that will receive later.

## 3.2 Functional Components

### 3.2.1 Log Collection

The log server uses the method of periodic storage. The rules are generated from the latest log and used to control in the latest data flow.

This system had taken two major methods to ensure the security and authenticity of log.

- (1) Identity authentication and tunnel technology are used.
- (2) The mechanism of syslog is used to transfer log.

### 3.2.2 Log Pretreatment

Log pretreatment [3-7] is an important process of data mining. The original information in log includes useless information and noise data which needs to do log pretreatment. The log pretreatment can make sure that the data is fit for data mining.

In the paper, the main pretreatment in system is changing 32-bit IP address to 8-bit binary value by HASH method.

### 3.2.3 Mining Association Rules

Mining association rules [8-10] is a method to find maximal mining frequent item set form data set. The rule reflects dependence between attributes.

If the association exists between attributes and we know one attributes then we can predict another. In this paper, apriori algorithm is used to generate mining association rules.

### 3.2.4 The Generation of Rule Table

Rule table is a rule set which generates by log pretreatment and data mining. It uses in packet filtering and is an important proof to generate black lists and white lists.

Rule table contains two main sources. One is generated by data mining. Another is generated by manger.

### 3.2.5 Audit Table

Audit tables contain catch table, white table and black table. After matching data flow with rule table, it generates dynamic access control mechanism by measuring the assessed value, updating rules and changing form one table to another.

By this way, system can measure the safety and speed up the conversion of packet processing speed.

### 3.2.6 Visualization

Visualization provides a friendly interface for manager. The main operations are CRUD.

## 3.3 AUDIT PROCESS

The process of the audit process is shown in figure 3.

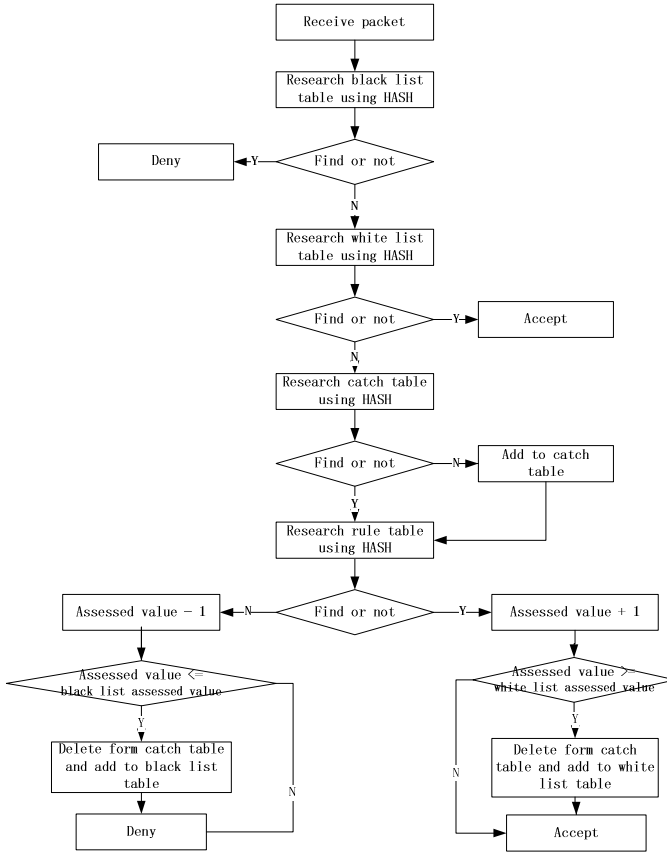


Fig. 3. The process of the security audit

## 4 Summary

This paper analyzed security audit based on data mining and realized dynamic access control. By the way of assessing, predicting and controlling the behavior of network, the system not only integrates static with dynamic but also integrate combine history with real-time.

**Acknowledgments.** I would like to express their thanks to Lin Weimin, Zhang Tao, Deng Song, Shi Jian, Wang Chen, and finally to my family.



## References

1. Yang, C., Liu, Q., Liu, L.: Network Security Audit Multi-level Model Brosa Audit. *Computer Engineering* 32(10), 159–160 (2006)
2. Zhao, P., Wang, H., Tan, C.: Design and Implementation of Network Isolation Security Audit System Based on Firewall Log. *Application Research of Computer* 24(7), 114–116 (2007)
3. Han, J., Kamber, M.: *Data Mining concepts and techniques*, pp. 30–51. China Machine Press, Beijing
4. Wang, R., Ma, D., Chen, C.: Institute of Surveying and Mapping, Information Engineering. *Application of Computer Technology* 69, 20–23 (2007)
5. Li, D.: Data mining technology and development trend. *Application of Computer Technology* 69, 38–40 (2007)
6. Huang, L.: The Application of Data Mining Theory in Security Audit Analysis. *Micro Computer Information* 23(93), 199–200 (2007)
7. Wang, G., Huang, L., Zhang, C., Xia, J.: ReView of Classification Algorithms in Data Mining. *Science Technology Review* 24(12), 73–77 (2006)
8. Wu, J., Shen, J., Wang, Y.: Improved Algorithm of Classification Based on Association. *Computer Engineering* 35(9), 63–66 (2009)
9. Deng, J.: Survey of Association Rule Data Mining. *Computer Learning* (3), 4–5 (2006)
10. Wang, Y., Jia, Y., Dong, X.: Data Mining Research Based on Association Rules. *Industrial Control Computer* 24(3), 86–87 (2011)

# A Study on Information Systems Risk Management Based on Control Self-assessment

Gengshen Yu<sup>1</sup> and Xiaohuan Zhu<sup>2</sup>

<sup>1</sup> Research Institute of Information & Communication Technology,  
State Grid Electric Power Research Institute, Nanjing, China

<sup>2</sup> Information Technology & Communication Company, NARI Group Corporation,  
State Grid Electric Power Research Institute, Nanjing, China  
yugengshen@sgepri.sgcc.com.cn, zhuxiaohuan@sgepri.sgcc.com.cn

**Abstract.** CSA (Control Self Assessment) evaluation as a new control method, is widely used in domestic and This thesis, d. through the study of control self assessment method in information system under the environment of the critical success factors and implementation process, explore the control self assessment on the risk management of information system improvement and the promotion effect, and evaluate control self assessment information system impact of risk management.

**Keywords:** Control self-assessment, CSA, risk management, information systems.

## 1 Introduction

With the development of enterprises, increasing concentration of information systems, business dependence on information systems continue to increase, the higher the level of information, the greater potential risk we take. How to better strengthen internal controls to protect information assets, security, integrity, true to ensure that information systems can effectively achieve business goals, must be in the process of information. Establishment of information systems auditing system and developing of information systems audit are essential to the process of information systems assurance and means. Traditional internal audit department is mainly concerned with the financial and business management and focus on books and data review. According to the "International Professional Practices Framework of internal audit," [1] Definition: "Internal audit is a confirmation and counseling. Designed to increase value and improve an organization's operations it through the system, standardized methods to evaluate and improve risk management, control and governance process is more effective to help the organization achieve its goals. "expanding the scope of internal audit to the level of risk control. "The future of the auditing profession and audit technology trends will come from the development of information systems audit", This view has gradually become a consensus of the auditing profession abroad.

Information systems risk management, internal audit has become an important part.

Information system of internal audit and risk management objectives of consistency. According to "internal audit specific criteria No. 28 - Information Systems Audit" definition "of information systems audit is through the implementation of information systems audit, information technology for the organization to achieve a comprehensive evaluation of management objectives, and make management recommendations based on evaluations, to assist information technology management organization to effectively meet their fiduciary responsibilities to achieve the organization's information technology management objectives. "risk management is to identify, control and eliminate or reduce the impact of IT system resources may not want the whole process time.

IT internal audit risk management is the new content. In the "Internal Audit International Professional Practices Framework" in the definition of internal audit is to improve the effectiveness of risk management to help them achieve their goals. IT-based information systems audit also have this goal.

Integration of IT technology, risk management, internal audit is an effective way. Risk management is a comprehensive work, the traditional risk assessment methods are not effective, comprehensive, in-depth risk assessment, integration of IT technology, internal audit through a top-down monitoring, evaluation, inspection, reports, recommendations to help comprehensive and effective risk management assessment, and risk management recommendations for improvement.

## 2 CSA

From the basic format, the, there are CSA namely:—conference guide method, questionnaire method management analysis [2].

Conference method is to guide the management and employees called up on a particular issue or course of conduct interviews and discussion methods. CSA guide Meetings Act there are four main forms: the control-based, process-based, risk-based and goal-based.

Questionnaire makes use of the questionnaire respondents as long as the tools to make a simple "yes / no" or "/" no" response, control program execution by the use of survey results to evaluate their internal control systems.

In management analysis, the management staff to learn business process layout. CSA guide (can be an internal audit staff) to the learning outcomes and their employees from other areas such as other managers and key staff to collect the information to be integrated. Through comprehensive analysis of these materials, CSA guide proposed a method of analysis, making the control program execution by CSA in their efforts to avail themselves of this analysis.

**Table 1.** A comparison between CSA and the traditional audit approach

<b>Project</b>	<b>Traditional auditing</b>	<b>CSA</b>
Form	By the auditors to assess and report	Self-assessment by the staff, all departments report
Drivers	Audit-oriented	Goal-oriented business
Scope and impact of	Policies, regulations, drive, external factors	Continuous improvement, learning curve of money, demand for internal improvements
Mode	Limited staff involved in	Broad participation and training of staff who
Form	Stakeholders concerned about the low level of	High degree of stakeholder concerns
Drivers	Through interviews, review of reports, records, etc. Control measures proposed by the auditors, suggestions for improvement	Through the research table, guide analysis to discuss the control structure Guided by the auditors, the low-level leaders or staff to make suggestions

CSA to early detection of risks more effectively to improve internal controls, increase employee understanding of organizational goals, understanding of risk and internal control, enhance operational and senior management are among the pass, control can reduce the cost.

### 3 CSA Application

#### 3.1 CSA Promotion

##### 3.1.1 To Obtain Management Support

The key to implemente CSA is the first level of the CSA to get encouragement and support, especially for the implementation of CSA's understanding of information systems, information systems and information systems of the CSA's role in promoting risk management; without their support, in the low-level management who would not in any serious implementation of the CSA process, there is no serious involvement attitudes, conduct CSA is a waste of time. Low-level managers also need to get the recognition and understanding, because the low-level managers most familiar with the situation, play a key role in the CSA. The main purpose is to eliminate fear and resentment, to explain the importance and effectiveness of control [4].

##### 3.1.2 CSA Method of Choice

Depending on the organizational culture, choose appropriate CSA method. CSA's success with the effective participation of employees is directly related to the domestic implementation of CSA may be more than the larger companies, geographical locations boast, with several branches, the use of hybrid of the CSA approach is more suitable. Combined with the results of questionnaire analysis and management, internal audit staff or department managers through regular surveys to analyze and evaluate the control effect.

##### 3.1.3 To Eliminate Resistance

Employee resistance mainly from two aspects, one workload, low-level implementation of CSA for the management of which is to be considered is the

increase in workload, the need to submit additional reports to management, which requires a guide CSA were to publicize the significance and high-level implementation of CSA work for the promotion; Second, critics fear stricken revenge, participating employees worried about the seminar, interviews, survey questions will be reprisals, afraid to speak their minds, so that the implementation of CSA effect on the greatly reduced, so the seminar, interviews, questionnaires need to make the appropriate state and the related security work, when necessary, can be anonymous manner.

### **3.1.4 Improve the Quality of Staff**

CSA is an important feature of the managers, internal auditors and other employees are involved in CSA assessment. Therefore, the quality level of these people will have a major impact on the effects of CSA. Improve the quality of personnel. Construction of a high-quality team, is to effectively play the role of internal control self-assessment is an important task; followed in the implementation of information systems control self-assessment process, CSA facilitator for information system risk management and information system control measures understanding of information systems, familiarity with laws and regulations directly affect the successful implementation of the CSA.

## **3.2 CSA Planning and Execution**

### **3.2.1 Identification Process, the Scope**

To determine which branch of the organization to implement CSA, select the appropriate scope of the CSA to help promote, through small-scale pilot, after the success of the promotion, you can select wide range of applications within the organization to pilot a system of information This will help the participants to the understanding of risk management information system to help CSA to promote the implementation within the organization; followed by the selection of participants and positions, the implementation of CSA does not mean that all personnel and positions need to participate, Select the key positions and are more familiar with the processes and operations personnel to effect the implementation of CSA played a role. CSA's overall goal is to service the organization's business goals.

### **3.2.2 Identify Risks, Assess Risks**

Identify the organization's risk, especially within the scope of CSA risk; information systems risk from the identifiable assets, vulnerability, threat of these three aspects to assess the sources of risk, and impact of the subject; in the implementation process also should consider audit risk, the International Auditing and Assurance Standards Board definition of audit risk = risk of material misstatement X detection risk, the assessment of the risks of material misstatement of the CSA to become the focus of the work of internal audit staff.

### **3.2.3 Identify the Control, Evaluation of Control**

To the establishment of the questionnaire can refer to the information related to standards such as ISO 27000 series, the corresponding national organizations in accordance with national requirements reference "GB / T 22239-2008 information

security technology to protect the basic level of information system security requirements", with the organization's key business processes and scope, the development of control questionnaire, and the control point empowerment emphasis.

**Table 2.** Control check list

No.	Control Objectives	Control Point	complete degree (1-10)	weight (1-5)
10	Strategic Planning	IT Value Management Assessment of current capabilities and performance IT Strategic Planning Planning methods		
11	Information Architecture	Continuous Improvement Enterprise information architecture model Confidentiality agreement Information Security Agency Staffing		
12	Process Framework	Authorization and approval Communication and cooperation Establishment of roles and responsibilities Information Security Coordination		
13	Management objectives and direction	Develop and maintain policies and procedures IT Strategic Plan IT organization and relationships		

**3.2.4 Report and Improvement**

Report is the culmination of the implementation of CSA, auditors need to report to senior management and the audit findings and recommendations of the audit report needs to be balanced, the report is to describe not only ineffective processes, controls and risk, also need to include effective control and has disposal of the risk, followed by even include process improvement, the proposed control measures and the implementation of the sequence and importance.

**4 Summary and Outlook**

Internal audit is an essential component of governance, can provide information on risk and control assurance, internal audit and information systems integration, contributes to improve the audit approach, improving information system controls that will enhance the ability of risk management information systems, development of organizational information system, risk and control understanding.

Control self-assessment is an ideal program which can be used by leaders to examine and improve their control systems Control self-assessment to deepen the organization's employees understand their risk and control, improve the ability to achieve their goals. Based on this, low-level leaders could manage values, vulnerability better and provide a good basis for risk decision-making [6].

## References

1. International Institute of Internal Auditors, International Internal Audit Professional Practices Framework (January 2009)
2. Li, H.: The Applied Research of the Enterprise Internal Control Self-assessment (CSA). Xiamen Science and Technology (4), 44–47 (2010) (Ch)
3. ISACA, CISA Review Manual 2010 (2010)
4. Zhu, H.: The CSA in the Environment of Information System Application. Productivity Research (22), 61, 93 (2008) (Ch)
5. Chen, S., Wang, Y., Xing, Q.: Internal Audit Role in Risk Management. Economic and Technological Cooperation (12), 82–83 (2010) (Ch)
6. Wang, S.: Risk Management is the Development Direction of the Internal Control. Economic and Technological Cooperation (12), 44–45 (2010) (Ch)

# An Implementation of User Interaction Effect by Extending jQuery in Web Application

Seon Ung Lee and Il-Young Moon

KUT

**Abstract.** In this paper, we propose an implementation of user interaction effect of web application using jQuery. jQuery is a very popular and extensible JavaScript library. It is strong in controlling DOM(Document Object Model), events and Ajax. But, it is poor at advanced effect of web user UI(User Interface). Implementing advanced effects need to extend jQuery or some skills. So, our proposal is a good way to implement UI effects.

**Keywords:** Web application, UI, jQuery.

## 1 Introduction

Web technologies are being developed as platforms. HTML5 is designed to be cross-platform and added new features like video, canvas and etc [1]. HTML5 takes three major technologies in broad sense. JavaScript is one of them that web developers must learn to specify the behavior of web pages [2].

Modern websites use JavaScript, especially Ajax (Asynchronous JavaScript and XML) and JSON (JavaScript Object Notation). It makes websites more dynamically and client-side scripting smaller. Nowadays, about a half of websites of the world uses jQuery that the best JavaScript library [3]. jQuery is simple and enables web developers to implement faster [4]. But jQuery too simple to implement advanced effect such changing background color smoothly or representing chained effect. There is a jQuery plugin specialized for UI, named jQuery UI. It's heavy and needs to customize the style of element.

We propose an implementation of user interaction effects of web applications. We describe it as implementing chained effect.

## 2 jQuery Basics

The jQuery philosophy is "Write less, do more." This philosophy can be further broken down into three concepts. First, Selecting elements and executing jQuery methods. Second one is chaining multiple jQuery methods on a set of elements. The other is using the jQuery wrapper and implicit iteration [5]. An example code of jQuery is below.



```
$('#div').hide();
```

The example means that selecting all div elements in DOM and making them be hidden. It corresponds to first and third philosophy of jQuery. jQuery methods for effect are show/hide, fadein/fadeout, animate/delay/stop, slideup/slidedown and toggle. These effect methods are queued implicatively each element of the set such as the multi thread.

But this system can't become aware of the priority of an effect. jQuery just can find out whether an element is animating from jQuery selector ':animated'. So the element in representing effect reveals the effect priority. Then the system can realize whether the effect can be canceled.

### 3 Revealing the Priority of Effects

The former chapter is described the jQuery basic and the problem of effect priority. So we should describe how to reveal the effect priority. We figure that there are two ways for revealing priority from our implementation. One is adding an attribute to animating elements. It uses attribute methods of jQuery which are 'attr' and 'removeAttr'. The other is extending the jQuery library. It is specifically making plugin or replacement of original function by new one.

#### 3.1 Using Attribute Methods

Using attribute as a variable, we need to know about WAI-ARIA (Web Accessibility Initiative – Accessible Rich Internet Application). It defines a way to make web content and web applications more accessible to people with disabilities [6]. It provides web authors with the description of roles and properties of widget presented. And it provides a framework for adding attributes to identify features for user interaction, and their current state. It is the reason why we need to know about WAI-ARIA.

In WAI-ARIA, 'aria-' prefix must be specified. For example, if a menu button is disabled, 'aria-disabled' attribute has 'true' value as following code.

```
<li role="menuitem" aria-disabled="true">menu1</li>
```

It is recommended to use 'aria-animate' attribute to describe the effect priority although it is not supported in WAI-ARIA document. The value type of the attribute is integer to show the level of the priority. The value might be from 1 to 10. It should be designed each system or website.

This method has some problems because attribute is in public. Other applications or frameworks can change the attribute. So, you use it carefully to prevent unexpected interruptions or effects. The best way to use attribute is to make the protocol among applications and framework interactions, especially Ajax live regions.

### 3.2 Extending the jQuery Library

This method is more programmable and reusable than above. But required knowledge is higher. Advanced JavaScript programming and jQuery structure should be known to extend library. jQuery structure is very familiar and ordinary in JavaScript. Tabs and Egyptian brackets are used here. The keyword ‘jQuery’ is equal to ‘\$’ returns a function [7]. The function has two parameter, selector and context. Table 1 contains core source of jQuery.

**Table 1.** Core source of jQuery

---

```

var jQuery = (function () {
  var jQuery = function ( selector, context ) {
    return new jQuery.fn.init( selector, context, rootjQuery );
  };

  // Map over jQuery in case of overwrite
  _jQuery = window.jQuery,

  // Map over the $ in case of overwrite
  _$ = window.$,

  // A central reference to the root jQuery(document)
  rootjQuery,

  [...]

  rootjQuery = jQuery(document);

  [...]

  return jQuery;
})();

```

---

From second to fifth line show us jQuery’s constructor code. When we call the jQuery functions, it creates and returns the “jQuery.fn.init” essence. jQuery is just a factory of “jQuery.fn.init” objects.

To make our jQuery plugins, we can expand the standard jQuery prototype. In the original jQuery code, “jQuery.prototype” is linked with “jQuery.fn”. Thus, following table 2 is the pattern of jQuery plugins. We implement this pattern to revealing effect priority.

**Table 2.** The pattern of jQuery plugins

---

```

new function (document, $, undefined) {

    var privateMethod = function () {
        // private method, used for plugin
    };

    $.fn.myPlugin = function () {

    };

    // and if you need a method that is not bound to the dom-elements:
    $.myPlugin = function () {

    };

}(document, jQuery);

```

---

## 4 Implementation

Our implementations take the jQuery plugin type because of reusability and variable scoping. As we announced previous chapter, every framework can modify the attribute value. JavaScript variable scoping enables jQuery plugin to hide the variable like the encapsulation of OOP (Object-Oriented Programming). The variable for saving priority of effect replaces the private method in Table 2.

The plugin we implement works like filter. It takes all of effect requests to the element. And it determines what to do after it checks the priority of requests. Furthermore, we also take the method that uses attribute because of web accessibility. It reveals the effect priority. When the framework asks an effect request, it will perceive how it works.

## 5 Conclusion

In this paper, we proposed an implementation of user interaction effect extending jQuery. It makes up for the problem in chained effect. We also introduce the way to make jQuery plugin. As jQuery is a popular and powerful JavaScript library, it's helpful for web developers. And web technologies are more developed, the accessibility of web application is important for web developers before designing UI. This field will be valuable for future web.

**Acknowledgments.** This work was supported by the Industrial Strategic technology development program (10040102, Development of self-initiated and hands-on e-Training service technology for the car mechanics) funded by the Ministry of Knowledge Economy(MKE, Korea).

## References

1. Mark, P.: HTML5: Up and Running – Preface, p. ix. O'Reilly (2010)
2. David, F.: JavaScript: The Definitive Guide: Activate Your Web Pages, pp. 1–3. O'Reilly (2011)
3. <http://trends.builtwith.com/JavaScript/>
4. <http://www.jquery.com/>
5. jQuery Community Experts. jQuery Cookbook, pp. 1–51. O'Reilly (2010)
6. <http://www.w3.org/WAI/intro/aria>
7. <http://www.splashnology.com/article/the-structure-of-jquery-dive-into-source-code/>

# An Intelligent Agent-Based Framework for Information Security Management

Chengzhi Jiang<sup>1</sup>, Bo Zhang<sup>2</sup>, Yong Yu<sup>1</sup>, and Xiaojian Zhang<sup>1</sup>

<sup>1</sup> Research Institute of IT and Telecommunication, State Grid Electric Power Research Institute

<sup>2</sup> IT and Telecommunication Company, State Grid Electric Power Research Institute  
{jiangchengzhi, zhangbo3, yuyong, zhangxiaojian}@sgepri.sgcc.com.cn

**Abstract.** Many organizations have introduced an Information Security Management System (ISMS) to improve their security information management. Contemporary management of information security involves various approaches, ranging from technology to organizational issues. According to international standard ISO/IEC 27001:2005, this paper presents an attempt to apply agent-based modeling to establish an information security management framework where machine learning technologies are incorporated. Within the framework, five types of agents are modeled in details. To foster the practical use of the framework in this paper, some critical issues are discussed.

**Keywords:** Information security management, agent-based modeling, intelligent agents.

## 1 Introduction

As a result of wider application of Information Technology in various business processes and growing spread of communication media, the need for information security is rapidly increasing to high levels for individual users and organizations. Information security assurance system may require a wide spectrum of knowledge, ranging from technical safeguard countermeasures to organizational security governance. Generally, information security requirements in enterprises may include two major inter-correlated categories: technical requirements such as firewall, data encryption, safe communication channel and managerial requirement such as security policy, organization of security and asset management. In terms of the security technologies used in enterprises, it is reported in CSI Computer Crime and Security Survey (2008) that nearly all organizations report the use of firewall and anti-virus software and 85 percent of organizations are using Virtual Private Network (VPN) [1].

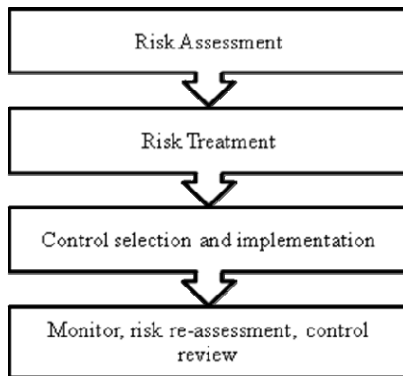
In this paper, we focus on the management of information security. With the aim of managing complicated processes of information security and providing high level information security, standardization studies have progressed throughout the worldwide management of enterprise information security [2]. ISO/IEC 27001 standard provides specification for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an Information Security Management System (ISMS) [3]. This standard is used worldwide by commercial organizations and governments as the basis for the management of organizational

policy and the implementation of information security [4]. The ISMS can be scaled in accordance with the need of the enterprise while implementing ISMS. With the increasing complexity of ISMS under uncertain environment, the underlying dynamics should be investigated to make ISMS more efficient and effective.

Agent technologies represent a new way of analyzing, designing and implementing complex systems. Agents can be used to model any scientific or behavioral phenomenon, organization, function to study the underlying dynamics of complex systems effectively [5]. A multi-agent model for implementing active security is proposed in [6] where a multi-agent Intrusion Detection System (MIDS) and an Active Security Mechanism (ASM) are modeled. An intelligent agent-based collaborative information security framework is proposed in [7]. In this model, the agents work with security system (IDS, Firewall, etc) and share attack and threat information between them. A Decision Support System (DSS) is developed in [8] to minimize false alarms made by IDS/IPS and decrease the response time to attacks. It can be seen that the application of agent technologies to information security often serve as the assistance to improve security technology such as firewall and IDS. In terms of managerial aspect of information security, factors that influence information security management are studied in [2] by means of surveys and statistical analysis. This motivates us to attempt to apply agents modeling in ISMS, trying to study its underlying complexity so as to improve its efficiency.

## 2 Agent Modeling

According to ISO/IEC 27001, risk management is the key process to the governance of organization and to protect the information security of the enterprise [4]. ISO/IEC 27005 provides the guidelines for information security risk management, supporting the ISMS that complies with ISO/IEC 27001 [9]. The risk management process is displayed in Figure 1.



**Fig. 1.** Risk management process

In this section, five types of agent are designed as Risk State Agent (RSA), Risk Assessment Agent (RAA), Risk Management Agent (RMA), Security Policy Agent (SPA) and Knowledge Base Agent (KBS).

## 2.1 Risk State Agent (RSA)

RSA stores the current risk states of assets relating to eleven control areas required by ISO/IEC 27001 standard. The attributes and methods of RSA are listed in Tab. 1.

**Table 1.** Structure of RSA

Attribute	Data type	Method
organization_id	Integer	cal_org_risk()
area_id	Integer	
area_weight	Double	
control_group_id	Integer	
risk_a	Double	
risk_b	Double	

It can be seen from Tab.1 that each instantiated RSA may represent the risk states of assets that are related to one control area (For example, Human Resource Security) with its weight. An example of mapping assets to corresponding control area can be found in the Information Security Risk Management framework proposed in [10]. For instance, the assets main groups for Human Resource Security may include IT staff, employee, users, contractors and owners. A group of controls under a specific area are identified by *control\_group\_id* with associated risk values before controls (*risk\_a*) and after controls (*risk\_b*). It is noted that under one control area, there might be a set of different combinations of controls. However, there is only one group is active, i.e. currently used. The method *cal\_org\_risk* is used to calculate the overall risk of the specified organization.

## 2.2 Risk Assessment Agent (RAA)

RAA is responsible for assessing and re-assessing the risks of assets relating to control areas. The attributes and method of RAA are listed in Tab.2.

**Table 2.** Structure of RAA

Attribute	Data type	Method
area_id	Integer	set_risk(area_id)
assess_period	Integer	
assets	List of Objects	
assets[i].threat	Double	
assets[i].impact	Double	
asset[i].risk	Double	

It can be seen from Tab. 2 that an instantiated RAA establish the connection with a RSA via *area\_id* that identifies the specific control area. *assess\_period* denotes how often the risks of assets are assessed and re-assessed. Each RAA comprises a list of asset objects, each of which has three attributes: *threat* that represents the likelihood of a threat exploiting the vulnerability of the asset; *impact* that denotes the impact of a

risk taking effect on the asset; *risk* that is calculated by multiplying *threat* and *impact*. The function *set\_risk* locates and updates risks of corresponding area via the parameter of *area\_id*.

### 2.3 Risk Management Agent (RMA)

RMA is in charge of the selection of implementation of risk controls. Its attributes and method are listed in Tab. 3.

**Table 3.** Structure of RMA

Attribute	Data type	Method
area_id	Integer	set_control(area_id, control_group_id)
risk_threshold	Double	
control_group_id	Integer	
controls	List of Objects	
controls[i].intensity	Integer	
controls[i].description	String	
cost	Double	

It can be seen from Tab. 3 that RMA is linked with RSA through two parameters of *area\_id*. *risk\_threshold* represents the acceptable risk level. Since the required controls in each control area can be fulfilled by various means, a group of controls may differ from another. Each group is composed of a set of control objects, each of which is stated by *intensity* that represents the intensity level of the control implemented and *description* that describes the control details. *cost* denotes the investment of the security controls. The method *set\_control* is used to update the *control\_group\_id* of the corresponding RSA.

### 2.4 ISMS Policy Agent (IPA)

ISMS policy is a set of rules that define the directions, objectives, principles and acceptable risk criteria with regards to the information security of an enterprise. The IPA in this paper carries the responsibility of setting parameters that guide the behaviors of other agents. Its attributes and method are listed in Tab. 4.

**Table 4.** Structure of IPA

Attribute	Data type	Method
area_id	List of Integer	
risk_threshold	List of Integer	
area_weight	List of Double	set_para(attributes)
assess_period	List of Integer	



According to ISO/IEC 27001, the candidate actions for treatment of risks include applying controls, knowingly and objectively accept risk, avoiding risk and transferring risks to other parties. The attribute of *risk\_threshold* is used to decide whether the RMA apply risk controls or accept the risk. The attribute of *area\_weight* of RSA will affect the calculation of the overall risk of the organization. By adjusting this attribute, IPA can make the overall risk reflect the key area that the enterprise pays extra attention on. In addition, IPA controls the risk assessment or re-assessment period via setting *assess\_period* of RAA. The method *set\_para* is used to adjust the corresponding parameters mentioned above. A possible setting for a specific control area is composed of {*area\_id*[*i*], *risk\_threshold*[*i*], *area\_weight*[*i*], *assess\_period*[*i*]}.

## 2.5 Knowledge Base Agent (KBA)

Learning in a multi-agent environment can help agents improve their performance [11]. KBA represents the past experience gained by implementing risk controls in the past periods. Its attributes and method are listed in Tab. 5.

**Table 5.** Structure of KBA

Attribute	Data type	Method
<i>area_id</i>	Integer	
<i>control_group_id</i>	List of Integer	
<i>risk_a</i>	List of Double	
<i>risk_b</i>	List of Double	

Case-based Reasoning (CBR) is able to make use of specific knowledge of previously experienced concentered problem situations (case) to solve a new problem situation by finding and reusing a similar case. Each instantiated KBA stores the cases (experience) of a control area identified by *area\_id*. It can be seen from Tab. 5 that each case in the knowledge (experience) base can be described as [*area\_id*, *control\_group\_id*, *risk\_a*, *risk\_b*] that can be learnt when selecting controls for a specific risk. It is noted that under one *area\_id*, there may be a set of available controls with associated effect of actions (difference between *risk\_a* and *risk\_b*).

## 3 Intelligent Agent-Based Framework

As the agents are defined respectively, the intelligent agent-based information security management framework is shown in Fig. 1. It is noted that the interactions between different types of agents are displayed.

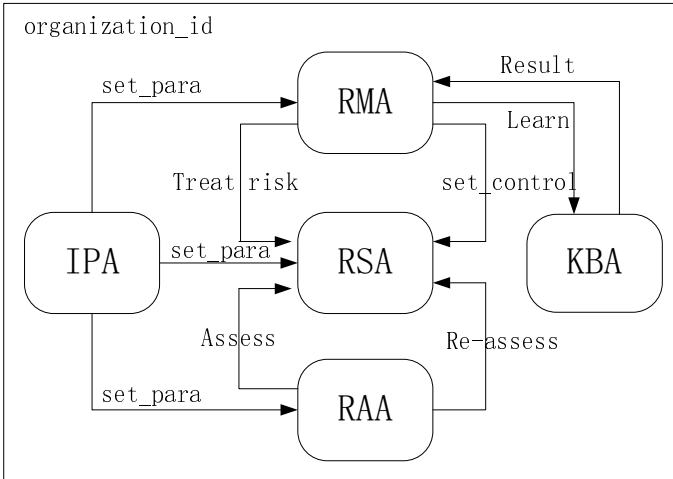


Fig. 1. Information security management framework

The workflow of the framework is described in details as follows:

*Step 1:* IPA sets the *risk\_threshold* of RMAs, *area\_weight* of RSAs and *assess\_period* of RAAs respectively.

*Step 2:* RAAs assess the risks of assets relating to eleven control areas and update *risk\_a* of RSAs.

*Step 3:* RMAs compare the *risk\_a* of RSAs to predefined *risk\_threshold* to decide whether to apply controls or not. If the current risk is acceptable, process is terminated. Otherwise, go to the next step.

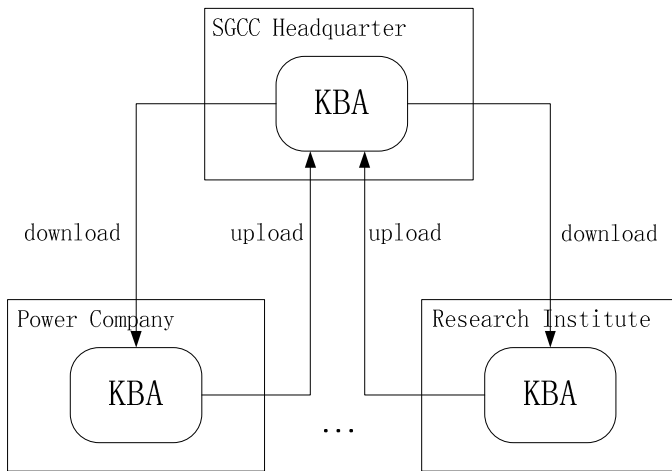
*Step 4:* RMAs inquire KBA for past similar cases for reusing or reference. If no similar case is found, a new case is created in KBA. Otherwise, the similar cases are input to RMA.

*Step 5:* RMAs apply the controls of similar case if there is any. Otherwise, RMAs will try controls that may achieve the expected security goal.

*Step 6:* RSAs may re-assess the RSAs in a predefined period in order to obtain the delayed effect of the corresponding controls applied. The *risk\_b* of RSAs is updated when the risk states are re-assessed.

Through iterative running of the processes, the KBAs will have more and more available cases (accumulated experience). Then, the learning process can be more useful.

It is suggested that the KBAs are shared in the organization having multiple hierarchies. Take State Grid Corporation of China (SGCC) for example, the implementation could look like the structure in Fig. 2. The subordinate power companies or research institutes can upload their cases to the KBAs of SGCC headquarter where the cases are merged, refined or filtered. The processed cases can also be downloaded from headquarter with appropriate access controls.



**Fig. 2.** Experience (KBA) sharing structure

## 4 Discussion

To foster the practical use of the proposed framework, it is suggested that the following points should be considered.

### 4.1 Risk Assessment

One of the most important tasks in risk assessment and re-assessment (Step 2 and Step 6 in the framework) is risk identification. Categorizing assets according to control areas defined in ISO/IEC 27001 may use the method proposed in [10]. It adopts STOPE (Strategy, Technology, Organization, People and Environment) view for the information security risk management. Both of the assets and security controls (eleven control areas in ISO/IEC 27001) are mapped to five domains of STOPE, thus making a mapping relationship between assets and security controls.

### 4.2 Learning Algorithms

The learning process of RMAs (Step 4 and Step 5 in the framework) basically follows the four steps of CBR that are retrieving the most relevant and similar cases from database of previous cases, reusing the experience and knowledge in retrieved cases to produce solution for new problem, revising the proposed solution and repair it if necessary and retaining the experience of the solution that may be useful for future problems. In the retrieval step, cases should be well-organized and indexed to facilitate case search and matching process. Reusing previous experience involves reviewing the past similar cases that may be successful or failed. Therefore, the actions to be taken should follow the right path and avoid the possible mistakes. The evaluation of the proposed solution for new problem is carried out in revising step by simulation or application in real-world environment. In the final retaining step, from a successful case, relevant problem descriptors and problem solutions can be stored as

learning resources. From a failed case, the failed conditions, solutions and repair records can also be stored to remind the system when a similar case is encountered in the future.

## 5 Conclusion

This paper presents an attempt to introduce agent-based modeling to research work of information security management. Base on the information security risks management requirement in ISO/IEC 27001 standard, five types of agents are modeled to carry out the tasks of risk assessment, risk controls as well as describing risk states and ISMS security policies. To improve agent performance, CBR that is a branch of machine learning technology is applied to risk management agent. When implementing the proposed framework, some critical issues such as risk identification process and learning algorithms are discussed. The future research may reside in enriching the framework to establish autonomous agents. That is, to reduce the manual intervention of the framework.

**Acknowledgments.** The research in this paper is supported by State Grid Information and Network Security Laboratory.

## References

1. CSI Computer Crime & Security Survey (2008), <http://gocsi.com/sites/default/files/uploads/CSISurvey2008.pdf>
2. Yildirima, E.Y., Akalpa, G., Aytac, S., Bayram, N.: Factors influencing information security management in SME: a case study. *International Journal of Information Management* 31, 360–365 (2011)
3. ISO/IEC-27001:2005. Information technology – security techniques – information security management systems – requirements
4. Humphreys, E.: Information security management standards: Compliance, governance and risk management. *Information Security Technical Report* 13, 247–255 (2008)
5. Govindu, R., Chinnam, R.B.: MASCF: A generic process centered methodological framework for analysis and design of multi-agent supply chain systems. *Computers & Industrial Engineering* 53, 584–609 (2007)
6. Zaki, M., Sobh, T.S.: A cooperative agent-based model for active security systems. *Journal of Network and Computer Applications* 27, 201–220 (2004)
7. Kuo, M.H.: An intelligent agent-based collaborative information security framework. *Expert Systems with Applications* 32, 585–598 (2007)
8. Kim, H.K., Im, K.H., Park, S.C.: DSS for computer security incident response applying CBR and collaborative response. *Expert Systems with Applications* 37, 852–870 (2010)
9. ISMS risk management. ISO/IEC 27005 (2008)
10. Saleh, M.S., Alfantookh, A.: A new comprehensive framework for enterprise information security risk management. *Applied Computing and Informatics* 9, 107–118 (2011)
11. Goldman, C.V., Renschein, J.S.: Mutually Supervised Learning in Multiagent Systems. In: Weiss, G., Sen, S. (eds.) *IJCAI-WS 1995*. LNCS, vol. 1042, pp. 85–96. Springer, Heidelberg (1996)

# Effect of Atmospheric Turbulence Coherence Degradation on the Scintillation Index for a Partially Coherent Laser Beam

Ruike Yang, Xiaoning Kang, and Xiaobo Zhang

School of Science, Xidian University, Xi'an, Shaanxi 710071, China  
ruikeyang@mail.xidian.edu.cn, xiaokang\_198668@126.com

**Abstract.** By using the theories of the mutual coherence function and the intensity scintillation of a laser beam propagating through turbulent atmosphere, the influence of light source partial coherence and the coherence degradation caused by turbulence on the scintillation index of a partially coherent laser beam are analyzed. The results show that atmospheric turbulence has a dual impact on the quality of a laser atmosphere communication channel. The intensity scintillation induced by turbulent atmosphere can increase the communication link bit error rate (BER), while the degeneration of coherence caused by atmospheric turbulence or adopting a partially coherent light source can decrease the BER within a certain distance range. Hence, the BER improvement by using a partially coherent laser as source may be restricted by turbulence coherence degradation for a laser atmosphere communication link.

**Keywords:** Text here free space optical communication, partially coherent beam, turbulent atmosphere, intensity scintillation.

## 1 Introduction

Free space optical (FSO) or laser atmosphere communication systems have the potential high data rate capability, low cost, quick deployment, and high security[1]. The optical intensity scintillation induced by turbulent atmosphere can cause the BER increase in the laser communication, especially, for strong turbulence. This can severely impair FSO communications systems quality. Due to the fact that partially coherent beams are less sensitive to atmospheric turbulence than fully coherent beams, one of the practicable methods is to use a partially coherent laser beam as the light source of communication system. It found that the partially coherence of the transmitted signal beam can improve the performance of laser communication systems and lead to a reduction in the BER of optical communication link[2,3,4]. In this paper, the coherence change caused by turbulence and partial coherent laser source with propagation distance are analysed. The influence of the turbulence degenerate coherence and light source partial coherence on the intensity scintillation are studied for a partially coherent Gaussian Schell model (GSM) beam propagating through turbulent atmosphere. It is significant to improve FSO communication link quality.

## 2 Coherence of a Partially Coherent Laser Beam in Turbulent Atmosphere

### 2.1 Spatial Coherence of Gaussian Beam in Turbulent Atmosphere

The field of a lowest-order Gaussian beam wave at  $z = 0$  in free space is defined by [5,6]

$$U_0(r,0) = a_0 \exp(-r^2/w_0^2 - ikr^2/2F_0) \tag{1}$$

where  $r = \sqrt{x^2 + y^2}$ ,  $w_0$  is the beam waist of the incident Gaussian beam,  $F_0$  is the phase front radius of curvature,  $\alpha_0 = 2/kw_0^2 + i/F_0$ . Based on Huygens-Fresnel integral the complex amplitude at  $z = L$  can be expressed as

$$U_0(r, L) = \frac{1}{p(L)} \exp(ikL) \exp(-\frac{r^2}{w^2} - \frac{ikr^2}{2F}) \tag{2}$$

where  $p(L) = 1 - L/F_0 + i2L/kw_0^2$ . The second-order moment, or the mutual coherence function (MCF), is defined by the ensemble average

$$\Gamma_2(r_1, r_2, L) = \langle U(r_1, L)U^*(r_2, L) \rangle \tag{3}$$

under weak fluctuation theory, it becomes

$$\Gamma_2(r_1, r_2, L) = U_0(r_1, L)U_0^*(r_2, L) \times \langle \exp[\psi(r_1, L) + \psi^*(r_2, L)] \rangle \tag{4}$$

The MCF takes the specific form[6]

$$\Gamma_2(r_1, r_2, L) = \Gamma_2^0(r_1, r_2, L) \exp[-4\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa) \{1 - \exp(\Lambda L \kappa^2 \xi^2 / k) J_0[|(1 - \bar{\Theta}\xi)p - 2i\Lambda\xi r|\kappa]\} d\kappa d\xi] \tag{5}$$

where  $I_0(x) = J_0(ix)$  is the modified Bessel function,  $p = r_1 - r_2$ ,  $r = (r_1 + r_2)/2$ ,  $\rho = |p|$ , and  $\text{Re}[\Delta(r_1, r_2, L)] = D(r_1, r_2, L)$  is the wave structure function (WSF) of the Gaussian-beam wave.

The MCF can also be used to predict the spatial coherence radius at the receiver plane. That is, the loss of spatial coherence of an initially coherence beam can be deduced from the modulus of the complex degree of coherence (DOC)

$$\begin{aligned} DOC(r_1, r_2, L) &= \frac{|\Gamma_2(r_1, r_2, L)|}{\sqrt{\Gamma_2(r_1, r_1, L)\Gamma_2(r_2, r_2, L)}} \\ &= \exp[-D(r_1, r_2, L)/2] \end{aligned} \tag{6}$$

The spatial coherence radius  $\rho_0$  is defined by the  $1/e$  point of the DOC. The WSF associated with a Gaussian-beam is defined by

$$D(\rho, L) = 8\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa) \exp(-\Lambda L \kappa^2 \xi^2 / k) \times \left\{ I_0(\Lambda \rho \xi \kappa) - J_0\left[\left(1 - \bar{\Theta} \xi\right) \kappa \rho\right] \right\} d\kappa d\xi \tag{7}$$

where  $\Phi(\kappa)$  is Von Karman spectrum i.e.,

$$\Phi(\kappa) = 0.033 C_n^2 \left( \kappa^2 + \frac{1}{L_0^2} \right)^{-11/6} \exp\left(-\frac{\kappa^2}{\kappa_m^2}\right) \tag{8}$$

The substitution of Eq. (7) into (6), the spatial coherence radius  $\rho_a$  of the Gaussian-beam under weak turbulence is expressed as

$$\rho_a = \left[ 8 / \left( 3(a + 0.62\Lambda^{11/6}) \right) \right]^{3/5} (1.46 C_n^2 k^2 L)^{-3/5} \tag{9}$$

$l_0 \ll \rho_a \ll L_0$

$$a = \begin{cases} (1 - \Theta^{8/3}) / (1 - \Theta) & \Theta \geq 0 \\ (1 + |\Theta|^{8/3}) / (1 - \Theta) & \Theta < 0 \end{cases} \tag{10}$$

Based on the method of effective beam parameters, by replacing the standard receiver beam parameters  $\Theta$  and  $\Lambda$  with a set of effective receiver beam parameters  $\Theta_e$  and  $\Lambda_e$ , the  $\rho_a$  of the Gaussian beam under strong turbulence can be obtained. For the Gaussian-beam, a set of effective receiver beam parameters is defined as

$$\Theta_e = 1 + \frac{L}{F_e} = \frac{\Theta - 0.81 \sigma_R^{12/5} \Lambda}{1 + 1.63 \sigma_R^{12/5} \Lambda} \tag{11}$$

$$\bar{\Theta}_e = 1 - \Theta_e, \Lambda_e = \frac{2L}{kw_e^2} = \frac{\Lambda}{1 + 1.63 \sigma_R^{12/5} \Lambda} \tag{12}$$

The effective spot radius is

$$w_e = w(1 + 1.625 \sigma_R^{12/5} \Lambda)^{1/2} \tag{13}$$

where  $\sigma_R^2 = 1.23 C_n^2 k^{7/6} L^{11/6}$ .

The output plane beam parameter is defined as

$$\Theta = 1 + \frac{L}{F} = \frac{\Theta_0}{\Theta_0^2 + \Lambda_0^2}, \quad \bar{\Theta} = 1 - \Theta \tag{14}$$

$$\Lambda = \frac{2L}{kw^2} = \frac{\Lambda_0}{\Theta_0^2 + \Lambda_0^2}, \quad w = w_0 (\Theta_0^2 + \Lambda_0^2)^{1/2} \tag{15}$$

The input plane beam parameter is defined as

$$\Theta_0 = 1 - \frac{L}{F_0}, \quad \Lambda_0 = \frac{2L}{kw_0^2} \tag{16}$$

### 2.2 GSM Spatial Coherence

If a diffuser is placed at the exit aperture of a coherence laser transmitter, the field of the optical wave emerging from the diffuser is[7,8]

$$\tilde{U}_0(r,0) = U_0(r,0) \exp[i\varphi(r)] \tag{17}$$

where  $U_0(r,0)$  is the field entering the diffuser and  $\varphi(r)$  is a random phase with zero mean. The correlation function associated with the random phase factor  $\exp[i\varphi(r)]$  of the diffuser can be described by a Gaussian function, ie.

$$\begin{aligned} B(r_1, r_2, 0) &= \langle \tilde{U}_0(r_1,0) \tilde{U}_0^*(r_2,0) \rangle \\ &= U_0(r_1,0) U_0^*(r_2,0) \exp\left(-|r_1 - r_2|^2 / 2\sigma_c^2\right) \end{aligned} \tag{18}$$

where  $\sigma_c^2$  describes the spatial coherence properties of the effective transmitted source.

To be consistent with results deduced from the GSM, a Gaussian spectrum[5] is used as follows

$$\Phi_s(\kappa) = \langle n_1^2 \rangle l_c^3 / (8\pi\sqrt{\pi}) \exp(-l_c^2 \kappa^2 / 4) \tag{19}$$

where  $\langle n_1^2 \rangle$  describes the index of refraction fluctuations associated with the phase screen and  $l_c$  is the lateral correlation radius.  $q_c$  is a nondimensional coherence parameter defined by  $q_c = L/(kl_c^2)$ , and  $l_c = \sqrt{2}\sigma_c$ . The larger the  $q_c$ , the worse the coherence performance.

By using the extended Huygens-Fresnel principle, the MCF at the pupil plane of the receiver for GSM is defined by

$$\begin{aligned} \Gamma_{pp,diff} &= \frac{w_0^2}{w^2(1+4\Lambda q_c)} \exp\left[\frac{ik}{L}\left(\frac{1-\Theta+4\Lambda q_c}{1+4\Lambda q_c}\right)r \cdot p\right] \\ &\times \exp\left[-(2r^2 + \rho^2 / 2) / (w^2(1+4\Lambda q_c))\right] \\ &\times \exp\left[-(\Theta^2 + \Lambda^2)\rho^2 / (1+4\Lambda q_c)l_c^2\right] \end{aligned} \tag{20}$$

The substitution of Eq. (17) into (7), the  $DOC_{pp,diff}$  of GSM in free space is become

$$DOC_{pp,diff}(\rho, L) = \exp\left[-\left(\frac{\Theta^2 + \Lambda^2}{1+4\Lambda q_c}\right)\left(\frac{\rho^2}{l_c^2}\right)\right] \tag{21}$$

Hence, we deduce the special coherence radius  $\rho_s$  caused by the partial coherence of light source

$$\rho_s = \sqrt{(l_c^2(1+4\Lambda q_c)) / (\Theta^2 + \Lambda^2)} \tag{22}$$

In practical application, atmospheric turbulence leads to the degradation of beam coherence when a laser beam wave propagating through turbulent atmosphere. The



spatial coherence of the partially coherent beam is influenced by turbulent atmosphere and the coherence degree of the light source in common. Therefore, the effective spatial coherence radius of the partially coherent beam in turbulent atmosphere is defined as

$$1/\rho_{eff}^2 = (1/\rho_s^2 + 1/\rho_a^2) \tag{23}$$

which describes the change of the coherence properties caused by turbulent atmosphere and the light source.

### 3 GSM Scintillation Index

Based on the beam effective parameter method, the Rytov approximation and Kolmogorov spectrum, the longitudinal component of the scintillation index in the presence of a diffuser which is restricted to weak fluctuations is defined by[9]

$$\sigma_{I,l,weak}^2(0, L) = \sigma_B^2 = 3.86\sigma_R^2 \times \left\{ 0.4[(1 + 2\Theta_{ed})^2 + 4\Lambda_{ed}^2]^{5/12} \cos[5/6 \times \tan^{-1}((1 + 2\Theta_{ed})/2\Lambda_{ed})] - 11/16 \Lambda_{ed}^{5/6} \right\} \tag{24}$$

Under strong fluctuation conditions, then we have

$$\begin{aligned} \sigma_{I,atm}^2(0, L) &= \sigma_{I,l,strong}^2(0, L) \\ &= \exp\left\{ 0.49\sigma_B^2/[1 + 0.56(1 + \Theta_{ed})\sigma_B^{12/5}]^{7/6} + 0.51\sigma_B^2/(1 + 0.69\sigma_B^{12/5})^{5/6} \right\} - 1 \end{aligned} \tag{25}$$

The radial component of the scintillation index is defined by

$$\begin{aligned} \sigma_{I,r,weak}^2(r, L) &= 4.42\sigma_R^2 \left( \Lambda / (1 + 4\Lambda q_c + 1.63\sigma_R^{12/5} \Lambda) \right)^{5/6} \\ &\times r^2 / \left( w^2 (1 + 4\Lambda q_c + 1.63\sigma_R^{12/5} \Lambda) \right) \end{aligned} \tag{26}$$

$$\begin{aligned} \sigma_{I,r,strong}^2(r, L) &= 4.42\sigma_R^2 \left( \Lambda_e / (1 + 4\Lambda_e q_c + 1.63\sigma_R^{12/5} \Lambda_e) \right)^{5/6} \\ &\times r^2 / \left( w_e^2 (1 + 4\Lambda_e q_c + 1.63\sigma_R^{12/5} \Lambda_e) \right) \end{aligned} \tag{27}$$

At the abaxial point in beam wave, the scintillation index can be expressed as the sum of two components

$$\sigma_I^2(r, L) = \sigma_{I,l}^2(0, L) + \sigma_{I,r}^2(r, L) \tag{28}$$

The effective parameters  $\Theta_{ed}$  and  $\Lambda_{ed}$  are defined by

$$\Theta_{ed} = 1 + \frac{L}{F_{ed}} = \frac{\Theta}{1 + 4\Lambda q_c} = \frac{\Theta_0}{\Theta_0^2 + \Lambda_0^2 N_s} \tag{29}$$

$$\Lambda_{ed} = \frac{\Lambda N_s}{1 + 4\Lambda q_c} = \frac{\Lambda_0 N_s}{\Theta_0^2 + \Lambda_0^2 N_s} \tag{30}$$

where  $N_s = 1 + 4q_c/\Lambda_0$ ,  $w_{ed} = w(1 + 4q_c \Lambda)^{1/2}$ .

### 4 Influence of Turbulence on the Coherence of a Partially Coherent Beam

Let  $\lambda = 1.55 \mu\text{m}$ ,  $w_0 = 2.5 \text{ cm}$ ,  $l_c = 0.1, 0.01, 0.001\text{m}$ . The relationships of the coherence radius of atmospheric turbulence or source with distance are shown as in Fig1. From Fig.1 we see that the coherence radius of source increases as propagation distance increases, the coherence radius of turbulence atmosphere decreases quickly as propagation distance increases. In fact, the spatial coherence of Gaussian-beam waves propagating through turbulent atmosphere depends on both turbulence atmosphere and light source.

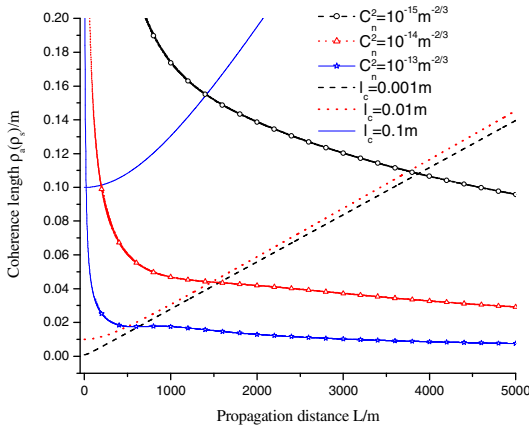


Fig. 1. The relationships of the coherence radius of atmospheric turbulence or source with distance

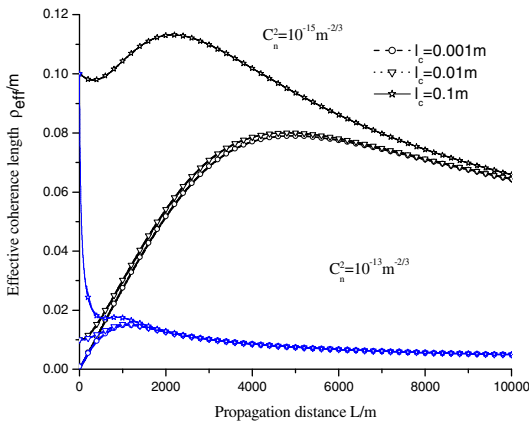
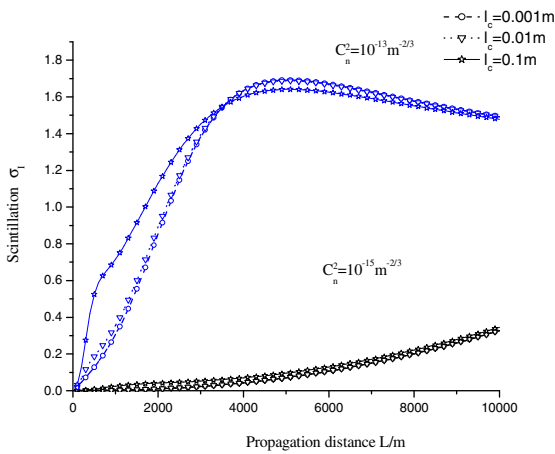


Fig. 2. The relationships of the effective coherence radius with distance

It is shown that the influence of the source coherence on the effective coherence radius decreases as propagation distance increases, the coherence radius of turbulence atmosphere decreases quickly as propagation distance. That is, due to the degradation of coherence induced by atmospheric turbulence, the coherence properties of a partially coherent beam decrease as the propagation distance increases. The better the coherence properties of source, the closer the distance acted by atmospheric turbulence. The effect of the change of source coherence on laser atmosphere communication exists only in a certain range.

## 5 The Scintillation Index of a Partial Coherent Beam Considering Turbulent Coherent Degradation

It is necessary to obtain the scintillation index for analysis of the BER performance of a partially coherent beam propagating in atmospheric turbulence. For a collimated GSM beam, the relationships between scintillation index and distance with different coherence radius  $l_c$  are shown as in Fig.3.



**Fig. 3.** Scintillation index of partial coherence beam with distance

From Fig.3, it is displayed that the scintillation index decreases as  $l_c$  decreases in weak and moderate turbulence. In strong turbulence, the scintillation index of a partially coherent beam is slightly higher than that of a fully coherent beam. The results indicate that the scintillation index depends almost on turbulence after propagating certain distance. That is, it is valid to use a partially coherent beam as source to reduce the scintillation index and the BER in a certain range.

## 6 Conclusion

From Fig 1. to Fig 3. for  $\rho_s < \rho_a$ , the intensity scintillation decrease depends on the coherence of source. However, for  $\rho_s > \rho_a$ , it almost depends on the degenerate coherence caused by atmosphere turbulence. The results indicate that using a partially coherent beam as source to reduce the scintillation index and the BER is valid in a certain range, and the coherence degradation caused by atmospheric turbulence need be considered.

**Acknowledgments.** The authors would like to thank the Support of the Advance Research Support Foundation of China.

## References

1. Ricklin, J.C., Hammel, S.M., Eaton, F.D., et al.: Atmospheric channel effects on free-space laser communication. *J. Opt. Fiber. Commun. Rep.* 3, 111–158 (2006)
2. Yang, R., Liu, Q., Wu, Z.: Effects of atmospheric turbulence scintillation on the error performance of partially coherent laser communication. In: *The Third European Conference on Antennas and Propagation Proceedings, EuCAP 2009*, pp. 1782–1785 (2009)
3. Ricklin, J.C., Davidson, F.M.: Bit error rate in a free-space laser communication system with a partially coherent signal beam. In: *Proceedings of SPIE*, vol. 4884, pp. 95–103 (2003)
4. Wang, S.C.H., Plonus, M.A.: Optical beam propagation for a partially coherent source in the turbulent atmosphere. *J. Opt. Soc. Am.* 69(9), 1297–1304 (1979)
5. Ho, T.L.: Coherence degradation of Gaussian beams in a turbulent atmosphere. *J. Opt. Soc. Am.* 60, 667–673 (1970)
6. Andrews, L.C., Miller, W.B., Ricklin, J.C.: Spatial coherence of a Gaussian-beam wave in weak and strong optical turbulence. *J. Opt. Soc. Am. A* 11(2), 1653–1660 (1994)
7. Andrews, L.C., Phillips, R.L.: *Laser Beam Propagation Through Random Media*. SPIE Press, Bellingham (2001)
8. Chen, X., Ji, X.: Influence of turbulence on spatial correlation properties of partially coherent annular beams. *Chinese J. Lasers* 36(9), 2319–2325 (2009)
9. Olga, K., Andrews, L.C., Phillips, R.L.: A model for a partially coherent Gaussian beam in atmospheric turbulence with application in Lasercom. *Opt. Eng.* 43(2), 330–341 (2004)

# The Application of Mobile-Learning in Collaborative Problem-Based Learning Environments

Chien Yun Dai<sup>1</sup>, Tzu-Wei Chen<sup>1</sup>, and Dar-Chin Rau<sup>2</sup>

<sup>1</sup> No. 129, Sec. 1, Heping E. Rd., Da'an Dist., Taipei City 106, Taiwan

<sup>2</sup> No. 111, Gongzhuang Rd., Zhonghe Dist., New Taipei City 235, Taiwan  
n98971108@gmail.com

**Abstract.** This article describes the design and delivery of a history course in visit historical tourism guild at elementary schools in Taiwan. The course was devised and deployed to supplement learning activities in the traditional classroom. A series of quasi-experiments was conducted with innovative instructional designs, that is, collaborative problem-based learning (PBL) environment, and their combinations. The impacts of these problem-based learning pedagogies on students, instructors, and course design were evaluated. The effects of problem-based learning pedagogies were mostly positive, thus reinforcing the instructor's confidence for further application to the rest of his courses. The authors further discuss the implications for schools, scholars, and teachers who plan to implement, or are already engaged in. In this study, the main goal was to improve students' learning, explore the problem-based learning effects, refine the history course, and reinforce the teacher's professional development. Therefore, mixed methods were applied. Specifically, the case study method was adopted to describe how students and the teacher adapted to the history course, and to detect deficiencies in design ideas and associated implementations for possible remedies in future trials. The authors intervened in students' learning via different instructional designs and deployed quasi-experiments to explore the effects of the intervention. In this study that is PBL class as the first step in our analysis, we used movie of history which about historical tourism guild, and change the classroom to visit historical tourism guild. The instructor designed and explained questions before visit. We need to observe and record students' behavior.

**Keywords:** Problem-based learning, collaborative learning, Mobile learning.

## 1 Introduction

Today, learners have grown up using devices like computers, mobile phones, and Tablet PC for almost every activity, from learning and work to entertainment and communication. This has altered the way they perceive and interact with the environment, both physically and socially. To chance the different cognitive requirements of the new concept, the educational community is considering new ways of learning. In particular, there is a wide interest in trying to engage learners with the appealing features of videogames and application tools.

One of the key aspects in collaborative learning is the implicit assumption that the learners from each other inside teams. Therefore, the way in which learners are grouped has a impact on the results of the learning process. Learning experience might turn into a negative one depending on the group composition. The coordination for group is other key aspect for the success of a collaborative learning experience. An effective way to coordinate group work is to assign concrete responsibilities to individuals using functional roles.

The grouping heterogeneous learners according to their learning style, we presume that the most autonomous learners will provide leadership and guidance to the group. At the same time, the effectiveness of the collaboration process within a team will improve by teaming up learners with complementary learning strategies and assigning them concrete interdependent responsibilities linked to a role. Teams are also a way to enrich social interaction among learners.

## **2 Literature Review**

### **2.1 Problem-Based Learning (PBL)**

Problem-Based Learning (PBL) can be explained as “the learning that results from the process of working towards the understanding or resolution of a problem”. Most cases, PBL is performed in small groups, fostering discussion and collaborative discovery, as the groups need to work with each other towards the solution for a specific problem or set of problems. A group has a tutorial leader or facilitator who shares information and guides the group through the learning process. In sum, PBL is a process of building new problem solving skills on prior knowledge by using critical thinking approaches and reflection (Maudsley, 1999).

Most of educational literatures have shown the benefits of using PBL and other approaches that promote active collaborative learning to improve learners’ thinking skills (Kreijns, Kirschner, & Jochems, 2003). It has been demonstrated that it leads to deeper levels of learning, critical thinking, shared understanding, and long-term retention of the learning material. Furthermore, collaborative learning also provides opportunities for developing social and communication skills, acquiring positive attitudes towards co-members and learning material, and building social relationships and group cohesion (Johnson & Johnson, 1994). However, the collaborative nature of PBL is frequently an issue. It is often difficult to arrange the schedules of all the members of a team to learners in work sessions. The meeting remotely can allow more effective PBL initiatives. New technologies have made to delocalize this approach, the so called dPBL (distributed Problem Based Learning) and sometimes CSCL techniques are used as virtual implementations of this pedagogical approach (Resta & Laferrière, 2007).

It is not easy to implement a PBL approach effectively when we factor out the direct personal concept. This is due to different reasons, and one of the most important ones is that PBL relies very heavily on group dynamics for its success. If group cooperation and cohesiveness are factors in PBL, that take into account that the lack of face-to-face interaction affects them negatively. Simply providing students

with some remote communication tools does not guarantee the emergence of the social interactions that lead to effective collaboration (Lurey & Raisinghani, 2001).

## 2.2 Collaborative Learning

The learning likes as a social activity within communities of practice, the better way to learn is with others in small group for learners to achieve a common goal (Bruffee, 1984; Chatti, Srirama, Kensche, & Cao, 2006). An active process may lead learners to deep develop the skills of critical thinking, communication, coordination and conscious knowledge construction mechanisms, called collaborative learning (Dillenbourg, 1999). Through interaction in collaboration, learning is not only to emphasize the individual accountability and the group interdependence, but also to achieve positive learning performance by inter-coordination between members (Chen, Chang, Lin, & Yu, 2008). Active exchange of ideas within small groups not only increase interesting and interaction among participators but also promotes critical thinking (Gokhale, 1995). With development of information technology, numerous Computer Supported Collaborative Learning (CSCL) environments were provided to assist learners engaged in collaborative learning activity. Cortez et al. (2004) provided a Mobile Computer Supported Collaborative Learning (MCSCSL) system to support school teachers through wirelessly networked computers and to promote students collaboration and constructivism with face-to-face contact. This study result showed that MSCSL system is able to enhance student's skill of new knowledge construction.

## 2.3 Interactivity (Context-Awareness)

Dey & Abowd (1999; 2004) defined "Context is any information that can be used to characterize the situation of an entity which is considered relevant to the interaction between user, application and around environment". Yang (2006) proposed an interactive model between learners/visitors and services thought context awareness. The services are provided with two dimensions:

- a) internal context such as ontext-user goals, tasks, communication, personal events, physical states, and so on (Prekop & Burnett, 2003);
- b) external context such Physical Environment and Technological Dimensions of Context (Lavoie, 2006).

Several context-aware applications that can provide users information in accordance with user's contexts have been developed. Chen, Li, and Chen (2007) proposed a personalized context-aware ubiquitous English vocabulary learning system which can provide appropriate content for learners in accordance with their location, leisure time and individual abilities. Moushiretal. (2007) provided a learning system which can suggests the best matched educational materials and peer helpers in accordance with the detected objects and the current location, and then allows the learners to share knowledge, interact, collaborate, and exchange individual experiences. Chen et al. (2008) developed a ubiquitous writing environment in which learners can use handheld devices to write in different situations through wireless networking technologies, handheld devices. Therefore, context-aware technology, various guided/learning environments can be built. In terms of kinds of systems, they can

bring different visiting/learning experiences for learners so that their degree of interest and learner can be promoted.

### 3 System Architecture

This study is to design and implement a collaborative problem-based learning system. The purpose of system is to enhance the learning interesting and motivation of learners for culture survey. The concept of design and development based on situated learning theory, and mobile device with context-aware technology to incorporate guided system of culture with situation of exploration of knowledge. By the interaction between learners, culture guided system and natural environment, learners are able to build and construct the knowledge of historical tourism guild that appears to the culture guide progressively as shown in Figure 1.and Figure 2 is system design

#### 3.1 System Architecture

This system is to design and implement a PBL system. The purpose of this system is to enhance the learning interesting and motivation of visitors for culture survey.

#### 3.2 System Construction

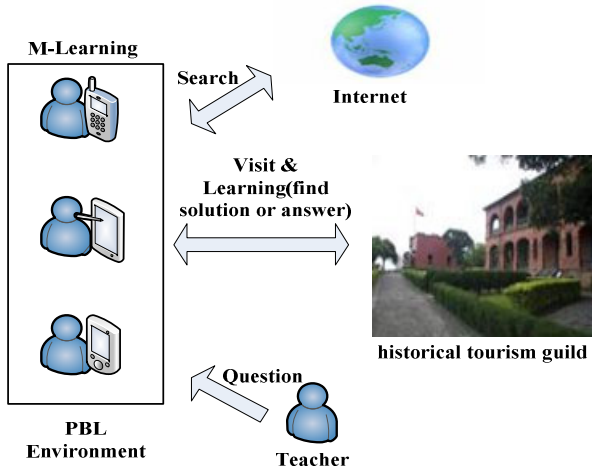


Fig. 1. System construction of collaborative problem-based learning environments

### 4 Conclusion

In this study, we have built a learning system to enhance visitor’s accomplishment of culture. The purpose of learning system is to extend learning opportunity for visitors who can enhance owns’ ability of knowledge sharing and accumulation through using handheld devices to visit, to discuss and create with other visitors in different situations. The experimental results showed that the learning system can effectively



promote visitor's learning performance and shared and accumulated knowledge, but it is unable to attract students' attention all the time. Therefore, we suggest some improved component in the future: a) Setup learning system into other devices with having larger screen sizes, such as Ipad; b) Adding phonetic input, let visitors be able to prompt record one's own experience and viewpoint; c) Adding exploratory activity of planning in advance, let visitors think the visit as playing games that can thus attract the participators' attention in cultural observation; d) Expanding the learning contexts of guide that should not only confine to a certain small area like as old Temple of Taipei city.

## References

1. Bruffee, K.A.: Collaborative Learning and the "Conversation of Mankind". *College English* 46(7), 635–652 (1984)
2. Chatti, M.A., Srirama, S., Kensche, D., Cao, Y.: Mobile Web Services for 3. Collaborative Learning. Paper presented at The Fourth IEEE International Workshop on Wireless, Mobile and Ubiquitous Technology in Education (2006)
3. Dillenbourg, P.: What do you mean by 'collaborative learning'? In: Dillenbourg, P. (ed.) *Collaborative-learning: Cognitive and Computational Approaches*, pp. 1–19. Elsevier, Oxford (1999)
4. Chen, T.-S., Chang, C.-S., Lin, J.-S., Yu, H.-L.: Context-Aware Writing in Ubiquitous Learning Environments. Paper presented at The 5th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), Beijing, China (2008)
5. Gokhale, A.A.: Collaborative Learning Enhances Critical Thinking. *Journal of Technology Education* 7(1), 22–30 (1995)
6. Dey, A.K., Abowd, G.D.: Towards a Better Understanding of Context and Context-Awareness: GVU Technical Report GIT-GVU-99-22, College of Computing, Georgia Institute of Technology (1999)
7. Dey, A.K., Abowd, G.D.: Support for the Adapting Applications and Interfaces to Context. In: Seffah, A., Javahery, H. (eds.) *Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces*, pp. 261–283. Wiley (2004)
8. Yang, S.J.H.: Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. *Educational Technology & Society* 9(1), 188–201 (2006)
9. Prekop, P., Burnett, M.: Activities, context and ubiquitous computing. *Computer Communications* 26, 1168–1176 (2003)
10. Lavoie, M.-C.: I, Mlearning: Identifying Design Recommendations for a context-aware mobile learning system. Paper presented at the IADIS International Conference Mobile Learning (2006), <http://www.mclavoie.com/projects/publications/LavoieIADIS06.pdf> (retrieved June 30, 2008)
11. Moushir, M., El-Bishouty, O.H., Yano, Y.: PERKAM: Personalized Knowledge Awareness Map for Computer Supported Ubiquitous Learning. *Educational Technology & Society* 10(3), 122–134 (2007)

12. Maudsley, G.: Do we all mean the same thing by “problem-based learning”? A review of the concepts and a formulation of the ground rules. *Academic Medicine* 74(2), 178–185 (1999)
13. Kreijns, K., Kirschner, P.A., Jochems, W.: Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior* 19(3), 335–353 (2003)
14. Johnson, D.W., Johnson, R.T.: *Learning together and alone: cooperative, competitive and individualistic learning*, 5th edn. Allyn & Bacon, Boston (1994)
15. Resta, P., Laferrière, T.: Technology in support of collaborative learning. *Educational Psychology Review* 19(1), 65–83 (2007)
16. Lurey, J.S., Raisinghani, M.S.: An empirical study of best practices in virtual teams. *Information & Management* 38(8), 523–544 (2001)

# Modeling Parallel MPI Programs in Petri Nets

Peng Zhang and Mei Qi

College of Information Science and Engineering, Shandong University of Science and Technology, 266510, Qing Dao Shan Dong, China  
bigbigroc@163.com, qmqimei@yahoo.com.cn

**Abstract.** The method of modeling concurrent programs is introduced. The basic framework of MPI programs are modeled by Petri net structures. On the one hand, the atomic actions of statements' execution are represented by the transition firing, on the other hand, properties such as Orphan Message, Lack of Message, Deadlock and Livelock Subnet of concurrent programs are analyzed and verified.

**Keywords:** Petri nets, Parallel Program, Verification, Deadlock.

## 1 Introduction

Parallel computing is the simultaneous use of multiple compute resources to solve a computational problem. To be run using multiple CPUs a problem is broken into discrete parts that can be solved concurrently. Each part is further broken down to a series of instructions, instructions from each part execute simultaneously on different CPUs. Message Passing Interface (MPI) is the most widely used message-passing system API and can be used to program parallel programs. One concept used in programming parallel programs is the future concept, where one part of a program promises to deliver a required datum to another part of a program at some future time. However programming under parallel environment is much more difficult than one CPU. Moreover, these CPUs often communicate each other, so we must ensure CPUS communicate correctly. Thus we need a verification tool to detect whether the program have errors of logic communication. Petri nets is a efficient tool in analyzing asynchronous parallel systems[1,2]. This paper gives the method on how to model MPI programs to Petri Nets(PN). With this method, the communication errors in parallel programming can be easily found.

## 2 Modeling MPI Programs

### 2.1 Introduction to MPI

Message-passing parallel programs is to divide a task into many processes, and one process take charge of part of work. One of common model is datum parallelisms. In this model, programs on each process is the same, but each deal with part of datum. SPMD (Single Process, Multiple Data) usually refers to message passing programming on distributed memory computer architectures. A distributed memory computer consists

of a collection of independent computers, called nodes. Each node starts its own program and communicates with other nodes by sending and receiving messages, calling send/receive routines for that purpose. Serial sections of the program are implemented by identical computation on all nodes rather than computing the result on one node and sending it to the others body.

### 2.2 Modeling MPI Programs into Petri Nets

In order to verification communication among CPUS and achieve less scale model, we concerned only for communication functions and basic flow of programs. We name a piece of program basic block that stand for sequence of instructions. We use a Tradition stand for block a Place in Petri nets stand for the execution condition of a block, Place with a token (dot in the Place) stand for the block can be executed.

The Modeling of general statements such as statements block, if statement and while statement is modeled as following.

### 2.3 Statements Block

The model of basic statements block is shown in FIG.1.  $t_1$  stand for sequence of instructions of a piece of statements.

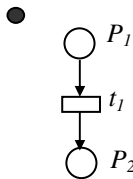


Fig. 1. Model of a basic block

### 2.4 If Statement

The syntax of the if statement is “if (expression) <statement>[else statement]”, for example , if  $a$  then  $b$  else  $c$ , the model of If Statement if shown in Fig. 2.

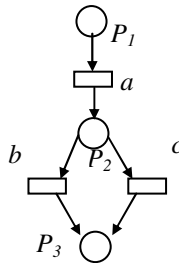


Fig. 2. Model of If Statement

### 2.5 While Statement

The syntax of the while statement is “while (expression)do < statement >” , for example ,while b do c ;d, d is the following statement. The model is shown in Fig. 3.

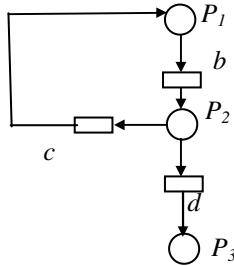


Fig. 3. Model of hile Statement

In MPI, There are over 100 library functions, all of the communication are realized by these functions. We mainly study the MPI program works in Point-to-Point blocking communication modes. The modeling of typical functions is as following.

### 2.6 MPI\_INIT and MPI\_FINALIZE

MPI\_INIT initializes the MPI execution environment and MPI\_FINALIZE cleans up all MPI state. Once this routine is called, no MPI routine (even MPI\_INIT) may be called. The model is shown as Fig4.

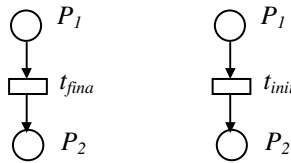


Fig. 4. Model of MPI\_INIT and MPI\_FINALIZE

### 2.7 MPI\_SEND(buf, count, datatype, dest, tag, comm)

MPI\_SEND may block until the message is received by the destination process. buf is the Initial address of send buffer (choice); dest is the rank of destination (integer) process. Tag is the message tag (integer). The model is shown as Fig 5.

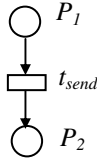


Fig. 5. Model of MPI\_SEND

**2.8 MPI\_RECV(buf, count, datatype, source, tag, comm, status)**

MPI\_RECV performs a standard-mode blocking receive from source process. Source is the rank of source process. The model is shown as Fig 6.

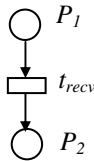


Fig. 6. Model of MPI\_RECV

**2.9 MPI\_SSEND**

Message send in synchronous mode. A send that uses the synchronous mode can be started whether or not a matching receive was posted. However, the send will complete successfully only if a matching receive is posted, and the receive operation has started to receive the message sent by the synchronous send. Thus, the completion of a synchronous send not only indicates that the send buffer can be reused, but also indicates that the receiver has reached a certain point in its execution, namely that it has started executing the matching receive. If both sends and receives are blocking operations then the use of the synchronous mode provides synchronous communication semantics: a communication does not complete at either end before both processes rendezvous at the communication. A send executed in this mode is non-local. The model is shown as Fig 7.

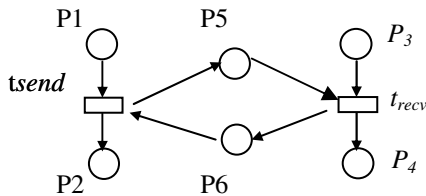


Fig. 7. Model of MPI\_SSEND

## 2.10 MPI\_BCAST

MPI\_BCAST broadcasts a message from the process with rank root to all processes of the group, itself included. It is called by all members of group using the same arguments for communication, On return, the contents of root's communication buffer has been copied to all processes. The model is shown as Fig 8.

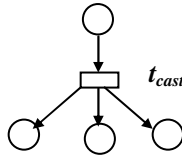


Fig. 8. Model of MPI\_BCAST

There are many other functions such as MPI\_GATHERV, MPI\_GATHER and so on, can be modeled in the same way. Because Petri nets and parallel program have similar properties, for examples deadlock, livelock. We can use reachability tree of Petri net to analyze the properties of parallel programs.

Example: The PN model of a MPI program is shown in Fig 9.

```
#include "mpi.h"
main(argc,argv)
int argc;
char **argv;
{
    char message1,message2;
    int myrank;
    MPI_Status  status;
    MPI_Init(&argc,&argv);
    strcpy(message1,"Hello world!")
    strcpy(message2,"This is MPI communication group!")
    MPI_Comm_rank(MPI_COMM_WORLD,&myrank):
```

```

If(myrank==0)
{
MPI_Send(message1,strlen(messag1),MPI_CHAR,1,99,MPI_COMM_WORLD),
MPI_Recv(message2,20,MPI_CHAR,1,99,MPI_COMM_WORLD,&status);
Processing1(message2);
Printf("Processed OK!:%s:\n",message2);
}
else if(myrank==1)
{
MPI_Recv(message1,20,MPI_CHAR,0,99,MPI_COMM_WORLD,&status);
Printf("received:%s:\n",message1);
Processing2(message1);
MPI_Send(message1,strlen(messag1),MPI_CHAR,3,88,MPI_COMM_WORLD),
MPI_Recv(message2,20,MPI_CHAR,3,88,MPI_COMM_WORLD,&status);
Processing2(message2);
MPI_Send(message,strlen(message2),MPI_CHAR,0,99,MPI_COMM_WORLD),
}
    else{
MPI_Recv(message2,20,MPI_CHAR,1,88,MPI_COMM_WORLD,&status);
Printf("received:%s:\n",message2);
Processing3(message2);
MPI_Send(message,strlen(message2),MPI_CHAR,1,88,MPI_COMM_WORLD),
}
MPI_Finalize();
}

```



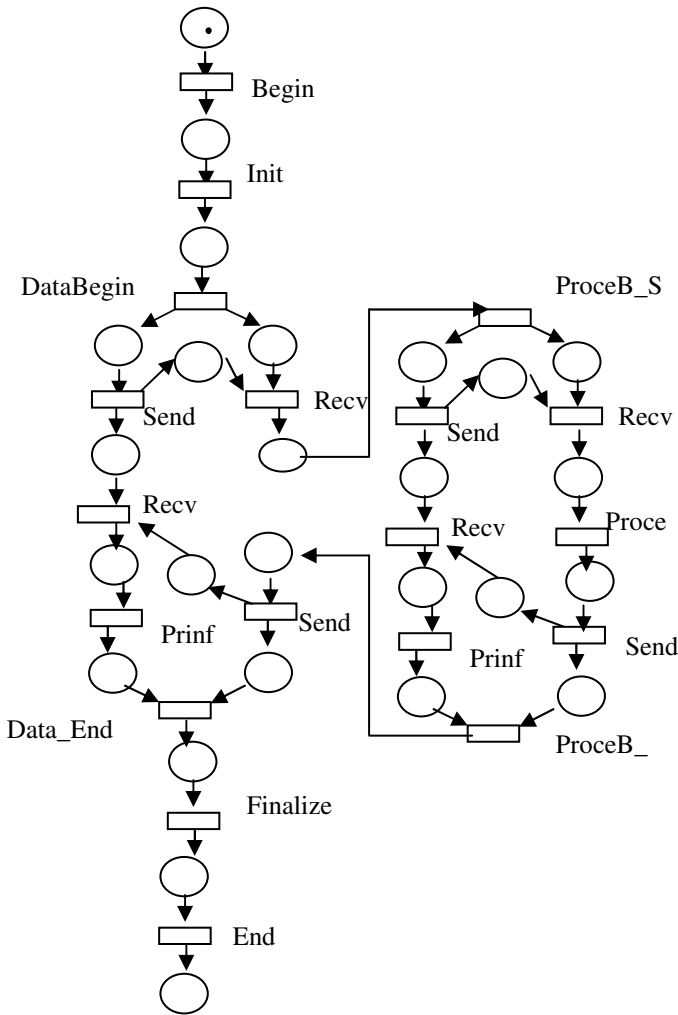


Fig. 9. Model of the MPI program

### 3 Conclusion

Petri Nets can be used to model parallel programs. If the code of a program is too long, the scale of the model is too large. We can use the top-down analyzing method to reduce it, that is, the stepwise refinement method[5]. First, give the outline model of the system, then consider the details of it, and finally get the total model. Then, with Petri net analyzing techniques and properties, verify whether the model has communication errors, such as deadlock, livelock, orphan message, and lack of message[6]. The other method is abstract analyzing. For a Petri Net model, with the refinement technique, abstract a subnet to a transition or place. It can greatly reduce the state of the model. Further study is suggested on the theory and method of automation programming verification.

**Acknowledgments.** Project is supported by National Natural Science Foundation of China(No. 60773034, 60803032, 60970001); Project supported by the Major Research plan of the National Natural Science Foundation of China(Grant No.91018007 ), Supported by the national Basic Research Program of Chian(973 Program) under Grant 2010CB328101. Supported by Research Project of SDUST Spring Bud(2009AZZ169). Peng Zhang College of Information Science and Engineering, Shandong University of Science and Technology, 266510, Qing Dao Shan Dong, China. bigbigroc@163.com, qmqimei@yahoo.com.cn.

## References

1. Jiang, C.: PN Machine Theory of Discrete Event Dynamic System, vol. 8. Publishing House of Science China, BeiJing (2000)
2. Murata, T.: Petri nets: properties analysis and applications. Proc. of the IEEE 4, 541–580 (1989)
3. Jiang, C.: Dynamic Invariance of Petri Net. Science in China 12, 567–573 (1997)
4. Peterson, J.: Petri net theory and the modeling of systems. Prentic-hall, Englewood Cliffs (1981)
5. Suzuki, I., Murata, T.: A Method for Stepwise Refinement and Abstraction of Petri Nets. Journal of Computer and System Sciences 27, 51–76 (1983)
6. Jiang, C., Lu, W.: On Properties of Concurrent System Based on Petri Net Language. Journal of Software 12, 513–520 (2001)

# The Design and Implementation of the IDS Catalogue Data Archive System

Li Jingang, Liu Shibin, and Liu Wei

Center for Earth Observation and Digital Earth, Chinese Academy of Sciences,  
Beijing, China

Lijingang.cn@gmail.com, sbliu@ceode.ac.cn, liuweixyr@126.com

**Abstract.** Catalogue Data Archive System (CAS), as a background system, is the core part of the Internet Data Service system for Remote Sensing Data (IDS), which is a web portal published by CEODE and provides remote sensing data sharing and purchasing services. CAS integrates a series of updating methods and technologies, such as a suitable data transmittal mode, reflection, flexible configuration files and log information in order to make the system with several desirable characteristics, such as ease of maintenance, open and scalable, stability and minimal complexity. This paper describes the four major components of the CAS (Data Collector module, Metadata Convert module, Data Archive module and Data Inject module) and some of the unique features of the system, technical problems encountered and resolved.

**Keywords:** Remote sensing data, archive system, C/S architecture, data management.

## 1 Introduction

IDS is a remote sensing data service system, which developed by CEODE. Its major responsibilities are data storage, catalog data retrieval service, product distribution service and free data download service. The data, managed by IDS, include satellite catalog data, standard product data and aeronautical data.

As the center of data management for IDS, the Catalogue Data Archive System (CAS) is responsible for improving the efficiency of data management, reducing the complexity of man-made operation, ensuring the whole system's reliable running and enhancing the level of services for the end-user. The CAS, as a back-end system, is the core part of IDS. It links the preceding and the following steps, acting as a key component. On the one hand, CAS needs to access the data exchange area which is a network RIAD to obtain product data handled by the data processing system. These data received from a series of satellites, such as SPOT1-5, LANDSAT-5 and LANDSAT-7 and so on. On the other hand, it is the data preparing stage for the following data distribution and application. Thus, the key tasks of the system are automatically receiving product data that need to be archived and constructing a product database and a product information database for online data discovery and retrieval services.

The system implements the C/ S architecture, which takes the database as the server and each module as a client. Those modules are loose-coupling, independent from each other, and could be configured flexibly, which contributes to the distributed management, maintainability, and expansibility of the system.

The purpose of this paper is to describe the design and implementation of those modules and particularly introduce the measures that have been taken to insure the integrity and validity of product data and flexibility and stability of the system.

## **2 The Core Design for CAS**

### **2.1 The Functions of the CAS**

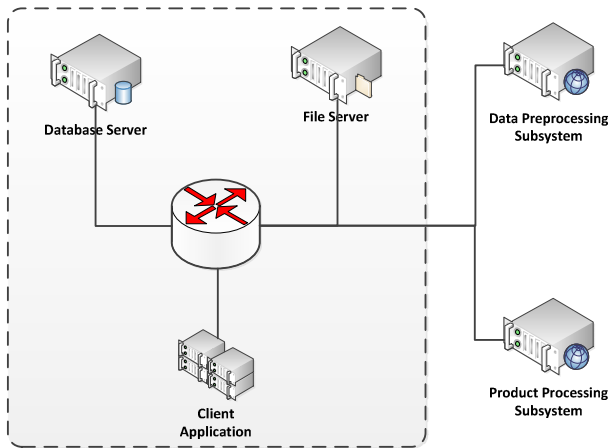
Due to many remote sensing data acquisition channels, each with different data format, data storage and data processing methods, the CAS must process different kind data in different way and storage them in unified format so that the IDS can provide a unified search and display mechanism.

The CAS accesses the predefined data exchange areas to obtain the data produced by the Data Preprocessing Subsystem and Product Processing Subsystem. After this, CAS extracts metadata information from these data and conforms to the given rules to store these remote sensing data and metadata in specific disk arrays, and then it generates JP2 picture and thumbnails for each remote sensing data. Finally, it imports the metadata information into the database. Thus, the CAS is composed of four modules, each of which implement part of the archiving workflow and executes independently. Furthermore, the CAS has the responsibility to maintain every step of archiving process and executing-state information stored in the database.

### **2.2 C/S Architecture**

C/S is the abbreviation of Client/Server. CAS takes C/S architecture in order to take advantage of the architecture's merits to enhance the whole system's flexibility and stability. Thanks to the architecture, the client can make use of the local resource sufficiently to reduce the load of the server and improve the interoperation. Moreover, the Point to Point transfer mode among the server and client can ensure data security and improve the response speed, which is advantageous for handling transactions with a huge amount of data [2].

The architecture is shown in figure 1. The CAS takes the Oracle database as the server end and the four modules as the client ends. These modules can be deployed on different client computers, which can make use of hardware resource of the client ends to handle a huge amount of remote sensing data. The server based on the Oracle database stores the product file metadata information and the state information referred to the CAS during the whole archival process. The centralized storing method can insure the coherence of information and reduce information redundancy.



**Fig. 1.** Hardware architecture of CAS

### 2.3 Database Design

As the server end, the database using Oracle 11g release 2 is mainly responsible for storing metadata information and implementing assistant functions to help every module running and interacting with each other. We designed three kinds of data tables, which are respectively called global configuration tables, archival workflow tables and metadata catalogue tables.

Global configuration tables records the entire configuration that all the modules share. For example, during the archive processing, every module is initialized by the storage configuration table which contains the network RAID information in term of the specialized storage ID as the primary key. Because of this smart design, the system's storage space can be expanded easily.

It is important for information preservation and exchange among these modules that the archival workflow tables provide brief information about product data, such as the whereabouts of the product data, archival time and so on. Meanwhile, the archival workflow tables report the logical marker that instructs the order of archival workflow to every remote sensing data. As the core task, the metadata catalogue tables contain product metadata information according to the data type.

### 2.4 Resuming Mechanism

During the archiving process, these modules record their real-time processing states, which are used for exchanging processing information among modules, by logic marks into the database. There are four kinds of logic marks represented by numbers. Number 0 represents the preparing state, number 1 represents the processing state, number 2 represents the successful state, and negative numbers represents the

unsuccessful state. Meanwhile, the negative numbers also represents the number of processing failures for every data.

In case the system dies for some unforeseen reasons, such as power-down, network issues and so on, these modules will automatically keep track of the previous states and continue to process the following steps according to the logic marks.

## 2.5 Reflection Mechanism

.NET Framework's Reflection API allows programmers to fetch type (assembly) information at runtime programmatically. We can also achieve late binding by using .NET Reflection. Reflection enables developers to use code that is not available at compile time. .NET Reflection allows an application to collect information about itself and also to manipulate on itself. It can be used effectively to find all types in an assembly and/or dynamically invoke methods in an assembly. This includes information about the type, properties, methods, and events of an object. With Reflection, we can dynamically create an instance of a type, bind the type to an existing object, or get the type from an existing object and invoke its methods or access its fields and properties.

## 3 The Four Modules

Figure 2 shows the CAS components and data/information flows. The main modules will be described in detail in this section.

### 3.1 Data Collector Module

Once the module waits for a period of time that we set, it can automatically retrieve information about all kinds of remote sensing data stored in the data exchange areas. The module depends on the XML configuration file to get the database information. Then, it queries the UNC paths for all the data exchange areas from storage configuration table. After that, it scans these disk arrays to get remote sensing data information and stores the latest data information into the ORIGIN IMAGE FILEINFO table and ORIGIN METADATA FILEINFO table. The reason to insert the data information into two tables is that some remote sensing data have two files: the metadata file and the image file, and store them independently can ensure to keep data error track. We construct this module to ensure the continuity of the archive operations.

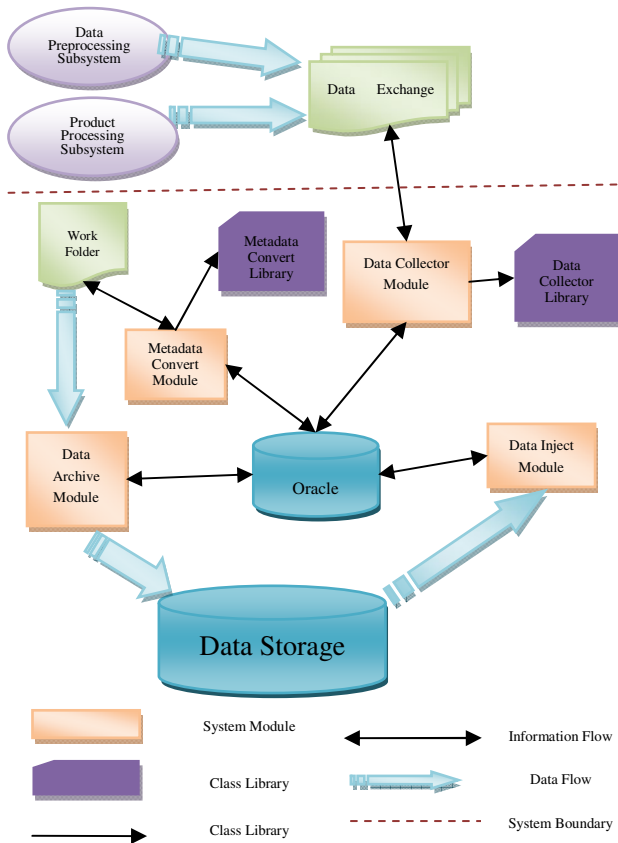


Fig. 2. Schematic representation of CAS components and data/information flows

### 3.2 Metadata Convertor Module

The product files of remote sensing data from different satellites are created in different formats. For the information about these files to be viewed through an online system, the files must be encoded in a more accessible format, such as XML format. XML is a perfect format for transferring data over the Internet that can facilitate the interaction of CAS with other companies and improve the system's scalability. Thus, this module invokes the methods in the class library to parse a file and extracts the metadata information from the product file. Finally, it encodes the metadata information into XML format. Nevertheless, different type of remote sensing data contains different metadata file type and different metadata information to be represented. According to the data format, we design special class to extract metadata information. Then on the basis of structure template that stores metadata items for every product data, this class can create metadata file for different type of remote sensing data. This metadata structure template is a XML format file that is stored in the global configuration table.

In order to display remote sensing data visually, CAS must create some images in different scales: 64×64, 128×128, 256×256, 512×512. This module makes these images after generated the metadata file. Finally, it generates a JP2 format image for the IDS system can display the data with geography coordinate in the world map.

### 3.3 Data Archive Module

The product data, the metadata file and its micro images are placed in a rigorously defined file management structure created by this module. The way to create the file management structure is based on the file name. According to the product file name rules, this module extracts relevant information from the file name, such as satellite name, sensor name, receiving time and so on, and then combines these elements into a canonical directory structure, namely, Satellite Name\ Sensor Name\ Receiving Time\.

There are two storage pools used to deposit these huge volumes of remote sensing data and each of the storage pools is about 10T capacity now. With the increase in the amount of data, we will put more storage pools in the system. How to automatically utilize and manage these pools to store files is a key problem which this module has to resolve. To handle this problem requires the following steps. Before depositing a file into a pool, this module queries a list of UNC paths from database about these available pools. Then, the module chooses one of the pools by order and checks whether there is enough capacity to store the file. If the capacity is enough, this module finally archives the file into the pool. Otherwise, the module sets a mark in the Storage Point Table to represent this pool is unavailable and chooses the next pool in turn. This step will go on until the module finds an available pool.

### 3.4 Data Inject Module

The function of the metadata catalogue module is to extract metadata information from metadata files and load the information into metadata catalogue tables. This information is very useful for searching for archived data online. This metadata information is divided into two kinds. One is general information, namely the data unconnected to spatial information, such as file name, sensor name and so on. The other is spatial information, such as the upper left latitude, the upper left longitude, the upper right latitude and so on. This information can be utilized for visualization of the remote sensing data. This module loads the metadata information using a very popular commercial database engine, ArcGIS Engine.

Because product data at different types are derived from the original data, their metadata shares a common part. For each type of the remote sensing data, the common part of the metadata is stored in a base table and the different parts are stored in an extended table that belongs to the specific remote sensing data. This module initialized by the metadata map configuration file which contains the metadata configuration information in term of the specialized satellite name as the parent node to decide which part inject to base table and which part to the specific extended table.



## **4 The Features of the CAS**

### **4.1 Data Transmittal Mode**

There are two categories of data transmittal mode. One is pull-mode; the other is push-mode. The CAS uses the former one. The files in the data exchange area are managed by the data producer, namely the pre-processing subsystem or processing subsystem are responsible for creating, deleting or modifying these files. Files stored in the data exchange area are preserved for a month and then will be deleted. As viewed from data consumer, namely the CAS, the data transmittal mode is pull-mode that means the data could be pulled from the two subsystems, data exchange areas, work folder to product data storage area. The virtues of the mode can reduce the complexity of the CAS and avoid the disoperations that lead the data loss.

### **4.2 Flexible Configuration File**

The system uses configuration files to implement configurability, expandability and maintainability. On the one hand, because the configuration files are separated from the program codes, just by modifying the configuration files, the system will fit to different conditions without reworking the code. On the other hand, these configuration files can be managed by a special program so that the modifying process will be easier.

The system uses two types of configuration files. One is XML configuration files each of which is used for a special module. The other is the global configuration table that is the common information stored in the database for all modules.

### **4.3 Log Information**

The CAS is a back-end system that has no user-interface. In order to keep track of the processing state of the system, it uses the log4net library to report the information. The log information is classified into different levels. For example, if you customize the Debug level, the log file will hide the other information that does not belong to the Debug level. This can enhance the flexibility of the system.

### **4.4 Expandability**

The system uses reflection to achieve late binding by using. When there are new remote sensing data type join in the system, we just implement new class based on the new data format to expand the system. Because of late binding, it is not necessary to modify the pre-existing code. We just amend the configuration files by which the modules initialized to decide which class to be instantiated to process the specific remote sensing data.

## 5 Summary

This paper introduces the architecture of the CAS, namely the C/S architecture, and its four modules. The key steps include the data access, metadata extraction, data archiving, and metadata catalogue. The CAS has a loose-coupling structure and leverages a series of technologies to implement stability, flexibility and scalability.

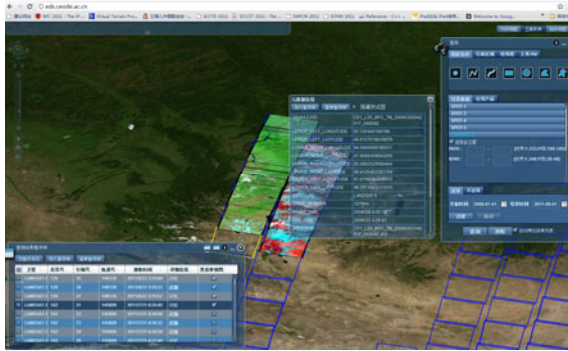


Fig. 3. The IDS system interface

The IDS system based on web service make it is easy that remote sensing data sharing. Through this platform users can get remote sensing data conveniently. The system interface show in fig.3. User can view data and operate on the map, such as pan, zoom and query. Besides, user can select data needed by pulling box on the map, and data service in platform can return data selected according to latitude and longitude. Finally, data selected can be overlaid on the map and be returned to the user.

**Acknowledgments.** I would like to express my gratitude to my supervisor, Mr. Liu Shibin, who has provided me with valuable guidance in every of the writing of this thesis. I shall extend my thanks to Liu Wei, without his help, I could not have completed my thesis successfully. At last, I'd like to thanks all members of the IDS project team for their encouragement and support.

## References

1. <http://eds.ceode.ac.cn>
2. Cui, T., Liu, S.-B., Liu, J.-B., Penh, H.-L.: A Monitoring and Archiving System of Remote Sensing Data Based on C/S. *Remote Sensing Information*, 59–62 (May 2008)
3. Liu, W., Liu, S.-B., Peng, H.-L.: The Design and Implementation of the HY-1B Product Archive System. In: *The 6th International Symposium on Digital Earth* (September 2009)
4. Li, M.-S., Saborowski, J., Nieschulz, J.: Web Service based spatial forest information system using an open source software spproach. *Journal of Forestry Research*, 89–90 (February 2007)

# Dem Generation with ALOS\_PRISM Stereo Data for Zijinshan Diggings

Xiaomei Zhang<sup>1</sup>, Guojin He<sup>2</sup>, and Chunxiang Chen<sup>3</sup>

<sup>1</sup> Center for Earth Obs. & Digital Earth, Grad. Univ. of Chinese Acad. of Sci., Beijing, China

<sup>2</sup> Center for Earth Obs. & Digital Earth, Beijing, China

<sup>3</sup> Zijin Mining Group Company Limited, Xiamen, China

**Abstract.** Digital Elevation Model (DEM) generation is one of the most important products of optical image data processing. The Japanese optical sensor ALOS/PRISM which was launched in January 2006 has stereo capabilities which can be used for a great number of applications and products. A cloud-free ALOS\_PRISM scene combination covering 35km x 35km Zijinshan diggings testfield has been analysed. It comprises the nadir and backward views with a base-to-height ratio of 1.0. The pixel size on the ground is 2.5m. The bundle orientation was executed with the related module of PCI Geomatica V10.3.2 software package and the RMSE(m) of 2.96 in X, 1.59 in Y were achieved in planimetry. This level of accuracy can be provided using the number of GCPs up to 16 which are distributed over the scene uniformly. Based on the scene orientation, a DEM of the area has been determined by an automatic image matching and PCI system yielded a DEM with 5m cell size. For the validation of extracted DEM, a group of GCPs selected over the testfield were utilized. In this analysis, GCPs were collected by GPS survey. This was done by ARCGIS 9.1 and mean square differences was obtained in the range of 3.93 to 12.5m. Moreover, A comparison between the extracted DEM of PRISM and ASTER GDEM(NASA) by using profile plot was conducted. In view of the overall trend, the two different curves of DEM were coincident. But The peak value for curve of ALOS DEM is more sharp due to the accuracy of ALOS DEM is more higher. In addition, ALOS DEM fits better where the variation of elevation is small. And the difference is more obviously where the variation is large. The Finally, 3 D views was generated using matched DEM and LANDSAT TM colour composite image in Zijinshan diggings area without problem. This was extremely useful in visualizing different parts of the area from different directions. And this is also proved to be of great help in identifying the tectonic geomorphology of Zijinshan diggings.

**Keywords:** ALOS\_PRISM, high resolution, DEM generation, validation, 3 D views.

## 1 Introduction

The Advanced Land Observing Satellite (ALOS) is a Japanese solution to high-resolution Earth observation. It is equipped with three mission instruments: Panchromatic Remote-Sensing Instrument for Stereo Mapping (PRISM), Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2), and Phased Array type L-band Synthetic Aperture Radar (PALSAR). The Panchromatic Remote-Sensing Instrument for Stereo

Mapping (PRISM) is a major instrument of ALOS. It has three independent catoptric systems for nadir, forward and backward looking to achieve along-track stereoscopy. Each telescope consists of three mirrors and several CCD detectors for push-broom scanning. The nadir-looking telescope provides 70 km width coverage; forward and backward telescopes provide 35 km width coverage each.

As shown in Fig.1, the telescopes are installed on both side of its optical bench with precise temperature control. Forward and backward telescopes are inclined  $\pm 24$  degrees from nadir to realize a base-to-height ratio of 1. PRISM's wide field of view (FOV) provides fully overlapped three-stereo (triplet) images (35 km width) without mechanical scanning or yaw steering of the satellite. Without this wide FOV, forward, nadir, and aft-looking images would not overlap each other due to the Earth's rotation.

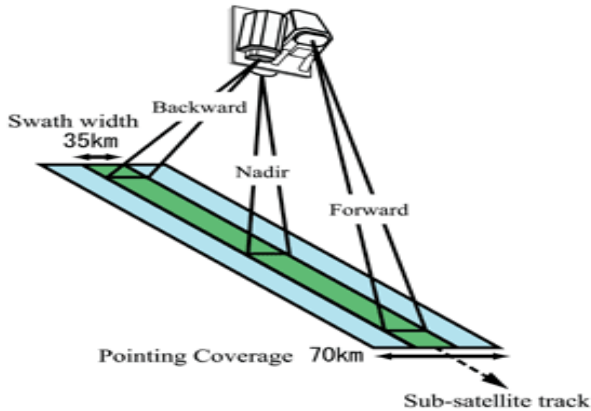


Fig. 1. PRISM Overview

PRISM Characteristics are shown in Table 1. (More details may be found at <http://www.restec.or.jp>)

Table 1. PRISM Characteristics

Item	Characteristics	Remarks
Number of Telescopes	3	
Wavelength	0.52 - 0.77 $\mu\text{m}$	
Base to Height Ratio	1.0	between fore and aft looking
IFOV	2.5 m	
Swath Width	70 km / 35 km	Nadir / fore and aft
S/N	$\geq 70$	
MTF	$\geq 0.2$	
Number of Detectors	28000 / band (Swath Width 70km) 14000 / band (Swath Width 35km)	
Number of Bands	1	Panchromatic
ADC	7bits	
Pointing	$-1.5^\circ$ to $1.5^\circ$	

NOTE: PRISM can't observe areas beyond 82 degrees south and north latitude.

The release of ALOS\_PRISM data has two significant impacts. First, they are available with very small expenses, compared with other high resolution(VHR) optical satellites capable of producing stereo images. Second, it provides a new alternative for mapping at large scales and for generating DEM. Especially, its along-track image acquisition configuration gives a strong advantage in terms of radiometric variations versus the multi-date stereodata acquisition with across-track stereo, which can then compensate for the weaker stereo geometry.

Up to till now, main emphasis has mainly given to the multispectral image application. Accounts of ALOS stereo-imagery being employed for DEM extraction are given in [4], [5], [6], and [7]. The First one studies deal with the accuracy estimates of ASTER stereoscenes with simulated data for DEM generation capability of the future planned satellite. Others report on an evaluation of the geo-positioning accuracy and DEM generation and validation from real ASTER data using commercial software packages. All these publications are out of china and not applicated into diggings, there are few reports about DEM extraction of ALOS\_PRISM in china.

In this paper, the authors report on the results of DEM generation approaches for ALOS stereo-scenes using PCI Geomatica V10.3.2 commercial software package. First, Zijin diggings testfield is described, then the application of bundle orientation is explained. This is followed by a discussion of DEM extraction using automatic image correlation. Afterwards, validation tests of matched DEM based on ASTER GDEM downloaded from NOVA web were given. Finally, 3 D views generation was carried out using the matched DEM and the Landsat TM data.

## 2 Pci Geomatica V10.3.2 System

The mathematical model which underlies and forms the basis of the photogrammetric solutions adopted in this package is based on the work developed by Toutin, T. [1-3] at the Canada centre for remote sensing (CCRS). It follows the threedimensional approach based on the use of collinearity equations which relates corresponding points in the image space and object space via the perspective centre of the imaging sensor. These equations have been adapted and formulated to suit the geometry of linear array (pushbroom) scanners such as ALOS\_PRISM in which each line of the scanner image has an individual and different perspective centre, instead of the single perspective centre for a whole image which exists with the frame photographs generated by aerial and space cameras. Besides the need to estimate and reconstruct the 3-D coordinates of the individual perspective centre for each individual line of a linear array image, it is also necessary to take account of the changing attitude of the satellite and its sensor over the time period during which the ALOS\_PRISM image has been acquired. Again, this is achieved through the modelling of the satellite orbital path in space by combining the satellite's positional and velocity vectors with the changing attitude of the platform to generate exterior orientation parameters for

the linear array image. Thus Toutin's model takes into account both the displacements due to the dynamically changing platform and sensor motion and orientation and those arising from the sensor geometry due to the physical characteristics of the Earth (rotation, curvature, and ground relief) (more details may be found at <http://www.pcigeomatics.com>).

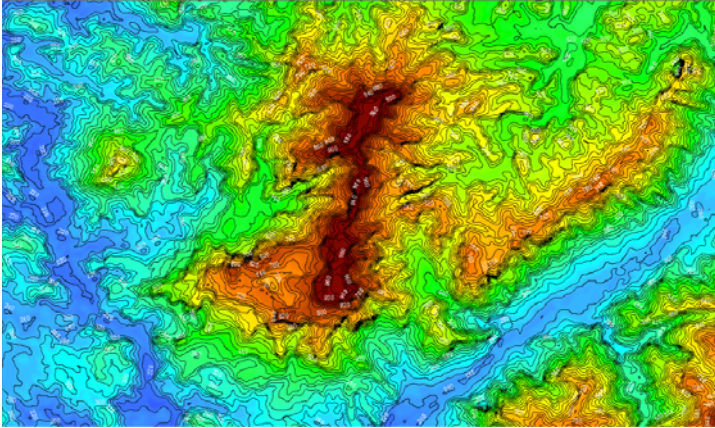
### 3 Zijinshan Diggings Testfield

The test site is Zijinshan diggings and its close vicinity, located in Shanghang country, Fujian Province, south-east China. It is famous for being China's largest gold output area. The elevation ranges from 200m to 1900m. The ZiJinshan diggings ALOS images are acquired on March 11th, 2010. The level 1B1 data (images, ephemeris and attitude) have been directly bought from CEODE of China. These images have a size of 4992x16000 x6 pixels for nadir and 4992x16000 x4 for back looking, both with a pixel resolution of 2.5m.

### 4 Dem Generation

To extract a DEM from a stereo-pair, it is necessary to match points on the one image with the corresponding points on the other image. For this purpose, PCI system employs an area-based image matching technique and produces the DEM through a comparison of the respective grey values on each of these images. This procedure utilizes a mean normalized cross-correlation matching method with a multi-scale strategy to match the image using the statistics collected in defined windows. Matching is performed by considering the neighbourhood surrounding a given pixel in the left quasiepipolar image (thus forming a template) and moving this template within a search area on the right epipolar image until a position is reached which gives the best match. The actual matching method employed with PCI software generates correlation coefficients between 0 and 1 for each match pixel, where 0 represents a total mismatch and 1 represents a perfect match. A second order surface is then fitted around the maximum correlation coefficients to find the match position to sub-pixel accuracy. The difference in location between the center of the template and the best matched pixel position gives the disparity or parallax arising from the terrain relief, from which the absolute elevation value is then computed.

This produces a regular grid of elevation values which are extracted to form the DEM. The interval between the points on the grid was 5m for stereo-model tested in this work. Fig. 2 shows the result of DEM extracted from ALOS\_PRISM, which was covered with its contours.



**Fig. 2.** DEM with contours in Zijinshan diggings

## 5 Dem Analysis

To go further and validate the data quality – more especially in terms of geometric accuracy – of the DEMs extracted from ALOS stereo-pair of Zijinshan diggings testfield, two different tests have been used:

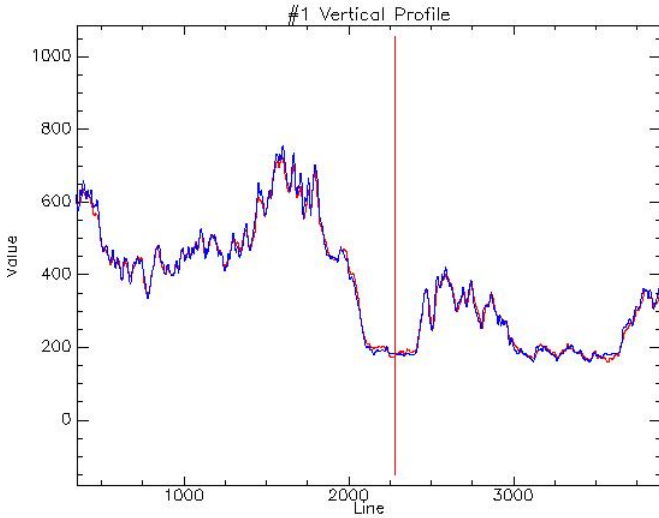
- (i) DEM accuracy check by GCPs measured by GPS;
- (ii) A comparison of the DEM of ASTER GDEM downloaded from the NASA web and the matched DEM generation with ALOS\_PRISM.

### 5.1 DEM Accuracy Check by GCPs Measured by GPS

Independent accuracy check of matched DEM was first carried out by GCPs measured by GPS survey. For this analysis, we collected 22 points and the software ARCGIS 9.1 was utilized. Then, we located the GCPs inside raster DEM in according to their planimetric coordinates and found their heights by interpolating the neighboured grids. For this, bilinear interpolation based on the neighboured 4 points has been implemented and RMSE-Z was in the range of 3.93 to 12.5m. This corresponds to a height accuracy of 2 – 5 pixels. All these values are based on the raw matching results, without any post-processing for blunder removal. This is the work we will do next.

### 5.2 Comparison of the ASTER GDEM against the ASTER DEM

For a further check of the data, a comparison between the extracted DEM of PRISM and ASTER GDEM(NASA) by using profile plot was conducted. The ASTER GDEM was downloaded from the NASA web and its resolution is 30 meter.



**Fig. 3.** Profile plots of ALOS DEM and ASTER GDEM (Left Xsection Right Ysection)  
 (Blue: ALOS DEM Red: ASTER GDEM)

From the figures, we can see that the similarity between two DEMs is quite visible. The peak value for curve of ALOS DEM is more sharp due to the resolution of ALOS DEM is higher. As a rule of thumb, the higher resolution can generate more detail. In addition, ALOS DEM fits better where the variation of elevation is small. And the difference is more obviously where the variation is large. (See the X section)

## 6 3 D Views GENERATION

We got a 3 D views derived from a LANDSAT TM colour composite image and the DEM from ALOS\_PRISM in the Zijinshan diggings area. 3D views draped on True colour composite have produced realistic terrain representations of the area and this was extremely useful in visualizing different parts of the area from different directions. This also proved to be of great help in identifying the textural variations associated with fault or linear alignment of the associated geomorphic units such as strath terrace. This type of variation is very difficult to observe in the traditional fieldwork. Fig. 3 shows the 3-D view of LANDSAT TM with DEM from ALOS\_PRISM.





**Fig. 4.** 3-D view of LANDSAT TM with DEM from ALOS\_PRISM

## 7 Conclusions

Based on ALOS stereo-scenes, with limited effort, a qualified digital elevation model was generated for the overlapped part of the images. For the check of accuracy, 22 control points have been collected by GPS survey, the horizontal accuracy that has been achieved within the range of 1 pixel-this is sufficient for mapping. We also compared the generated DEM from PRISM with the DEM generated from ASTER GDEM. The trend of two DEMs are similarity except that the peak value of the ALOS DEM is more sharp, this is because the higher resolution can generate more detail. From the results we think the accuracy can satisfy the requirement of DEM generation and line mapping with PRISM imagery. If we compare these matching results with those which were obtained earlier with other satellite sensors of similar type (SPOT-5, IRS-P5) we note that the accuracy (expressed in pixels) is about the same as with IRS-P5, but less good than SPOT-5.

**Acknowledgments.** This work was carried out under the National Natural Science Foundation of China (NSFC) under Grant No. 60972142.

## References

1. Toutin, T.: Multi-Source Data Fusion with an Integrated and Unified Geometric Modelling. *EARSEL – Advances in Remote Sensing* 4(2), 118–129 (1995)
2. Toutin, T., Cheng, P.: DEM generation with ASTER stereo data. *Earth Observation Magazine* 10, 10–13 (2001)
3. Toutin, T.: DEM Extraction from High Resolution Imagery. *Geospatial Today* 2(3), 31–36 (2003)
4. Chen, T., Shibasaki, R., Shinohara, F., Wang, C.: DEM Generation With Prism, Alos Simulation Data A Case Study using Air-Borne Three Line Scanner Imagery. In: *Asian Conference on Remote Sensing (ACRS 2005)*, Hanoi, Vietnam (2005)

5. Marangoz, A., Büyüksalih, G., Büyüksalih, I., Sefercik, U.: Geometric Evaluation, Automated DEM and Orthoimage Generation from Along-Track Stereo ASTER Images. In: Geometric Evaluation. 2nd International Conference on Recent Advances in Space Technologies Space in the Service of Society, RAST 2005, Istanbul, Turkey, June 9-11 (2005)
6. Wolff, K., Gruen, A.: DEM generation from early ALOS/PRISM Data using SAT-PP. In: High Resolution Earth Imaging for Geospatial Information on ISPRS Hannover Workshop 2007, Hannover, Germany, May 29- June 1 (2007)
7. Bignone, F., Umakawa, H.: Assessment of ALOS PRISM Digital Elevation Model Extraction over Japan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37, 1135–1138 (2008)

# A Novel Smart Home System

Ligeng Yu<sup>1</sup>, Xueen Li<sup>1</sup>, and Weijuan Han<sup>2</sup>

<sup>1</sup>The Hi-tech Innovation Engineering Center,  
Institute of Automation Chinese Academy of Sciences

<sup>2</sup>School of Mechanical Electronic and Information Engineering,  
China University of Mining and Technology, Beijing

{ligeng.yu,xueen.li}@ia.ac.cn, hanweijuan\_86@126.com

**Abstract.** Electronics such as air-conditioner, TV, humidifier, hi-fi equipment, can be called energy management systems implementing monitoring and statistics of energy data in the building. Their purpose is to make the relationship between power consumption and comfortable and healthy life environment more reasonable. In this paper, we present a novel smart home system integrating many kinds of energy efficiency features. This system is called VCE, which means making healthy life easier. VCE has many distributed low-power sensors in the environment which is used for detecting and better regulating human environment. These data are used for monitoring and analyzing real-time consumed energy on device level. VCE can adjust the temperature, humidity, light intensity volume and other information. And VCE can also detect human real-time vital signs like blood pressure, through sophisticated sensors.

**Keywords:** Smart home system, low-power sensors, healthy environment.

## 1 Introduction

Most of the energy management systems only implement monitoring and statistics of energy data which is collected from smart meter or consumer electronics [1]. Under these energy management systems, people should manually control each device to reduce energy consumption and create comfortable environment.

Intelligent home control system takes home control network as its basis, forms an information interaction platform of intelligent household appliances, receives and judges external states and commands through the intelligent home information platform and realizes various functions of intelligent home on this basis, thereby making householders enjoy an intelligent and humanized home life[2].

Several efforts have addressed the development of a smart environment system through networking existing devices and resolving interoperability issues with the help of middleware [3]. They are NGN@HomeETSI[4], TEAHA[5], EPERSPACE [6], AMIGO[7], ATRACO[9], HOME2015[10] SENSE[8]and so on. The purpose of them is to present solutions based on communication between electronic machines to solve the interoperability problem in smart home environments.

Main concept of VCE is making relationship between power consumption and comfortable healthy life environment more reasonable, in a sustainable way. It also

integrates various technologies, products, services, from consumer appliances and equipment using advanced sensing, communications, and control technologies. Sensing oriented: the home gateway gathers sensing data and special event information from the personal area network deployed in home environments. These sensing units provide information to the decision maker which is home gateway or people. The collection of information as the following: its environmental attributes (e.g., temperature, noise level, light intensity), and the people, physical objects, and computing entities that it contains. Healthy service oriented: The decision component then adaptively selects the correct home services based on the current home state of affairs and personal health status. VCE attaches great importance to the feedback of people's different health status on the environmental factors.

## 2 The Architecture of VCE

The implementation of VCE integrates zigbee network, mobile network, web internet in terms of home gateway and smart terminals.

The following Fig1 depicts the network architecture of VCE. In a quick view the VCE is comprised of different sensing components running in different location of the network, such as mobile phones as clients and PC, home gateway as Server and software for accessing various home devices like automation equipments and network cameras. Smart terminal is not to make people's lives lazier, but healthier and more comfortable. The smart terminal is not super virtual entertainment system, but efficient energy management system. The depicted network architecture shows that the home user can have control over his house either from inside using a PC or a mobile phone. Remote access is provided through a dedicated portal server after the user has authenticated him by giving a username and password. The authentication procedure applies on both GUI clients for mobile phone and remote PC.

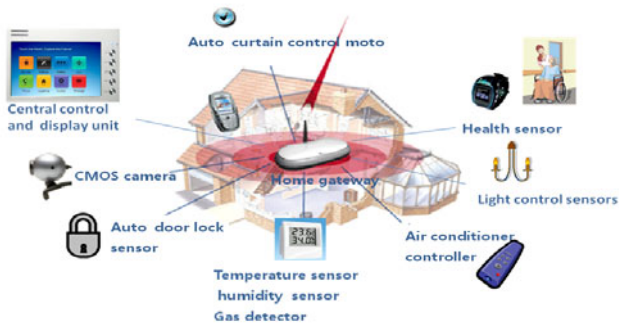


Fig. 1. The architecture of VCE

The connections on the Portal server are encrypted using strong encryption algorithms while the connection between the Home Gateway and the Portal is initiated from the Gateway and is encrypted as well. This mechanism on one hand provides the user with advanced protection from harmful internet attacks and on the

other hand offers simple and easy installation since it doesn't need any network configuration on the modem/router.

### 3 Topology of VCE

We use 433M-10k fsk-modulation transmitter. By adjusting the transmit power, we can change the communication distance between home gateway and smart terminals. The max penetration distance is three floors. Given the max electromagnetic penetration distance, star network is the most suitable network topology. The entire network adopts 802.15.4 GTS synchronous communication mechanism as is shown in Fig2. During CAP's free competition period, smart terminal gets its communication time slot of home gateway. After obtaining time slot, terminal can communicate with home gateway at its time slot during CFP phase. Smart terminal works on the basis of sense / sleep rotation mechanism, reducing power consumption.

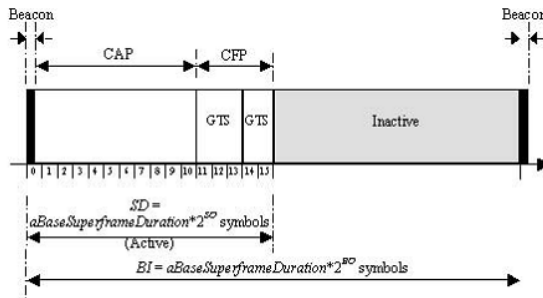
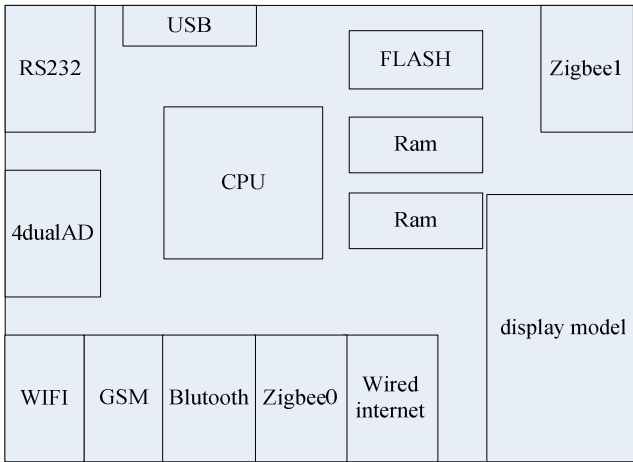


Fig. 2. GTS mechanism of 802.15.4

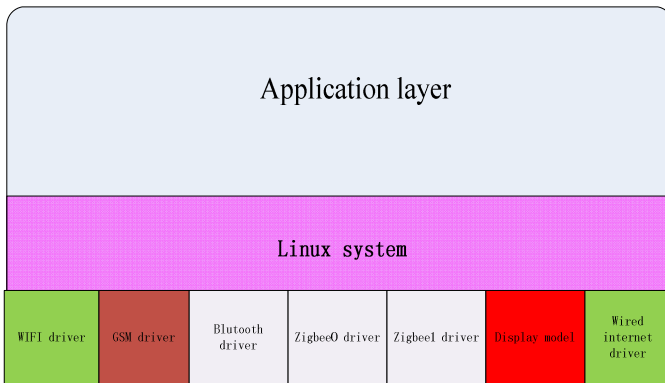
### 4 Home Gateway

Home gateway is the heart of intelligent home control system, through which to achieve information collection and centralized control. Its functions generally include home security, video intercom, remote meter reading, appliance control, home information services, value-added services. Network gateway device design concept is: independent calculation and display platform, free of pc system. Gateway is very important form of organizing energy.

The schematic diagram of home gateway is shown as Fig3 and Fig4. VCE's home gateway employs Linux operating system which has many scalable interfaces. Microprocessor S3C2440A including ARM9 core is selected as main CPU. Home gateway needs to save a large number of data, so 64MB FLASH and 128MB sdram is used to save system configuration. Special power monitor chip is used for power management. The communication chips include wifi GSM Bluetooth and zigbee.



**Fig. 3.** The hardware configuration of home gateway



**Fig. 4.** The architecture of home gateway software

Home gateway fused with many kinds of computer techniques, and therefore it has its own unique characteristics. The characteristics of home gateway mainly include:

- 1) High reliability, the software and hardware running on it must be very reliable; [2]
- 2) Low power consumption, home gateway provides a 24-hour uninterrupted Internet broadband access, so energy consumption must be decreased to the minimum;
- 3) There must be a secure environment which can ensure the operation of various Internet services;
- 4) It has the ability of short-distance communication such as zigbee, Bluetooth and remote communication like wifi, wired Ethernet.
- 5) It can fuse sensing data and make decision. For instance, when the temperature is blew 20 centigrade which is expected to be 25 centigrade. Then home gateway

should send command to air conditioner “warm the house to 25centigrade”. When the blood pressure of old people is abnormal, home gateway should adjust environment characteristics, for example light sound temperature and humidity. At the same time home gateway should send GSM message to the patient’s family. So home gateway is the core device that makes smart terminal connection to each other in the form of fusing sensing data.

6) It must be scalable because of the rapid developing technology.

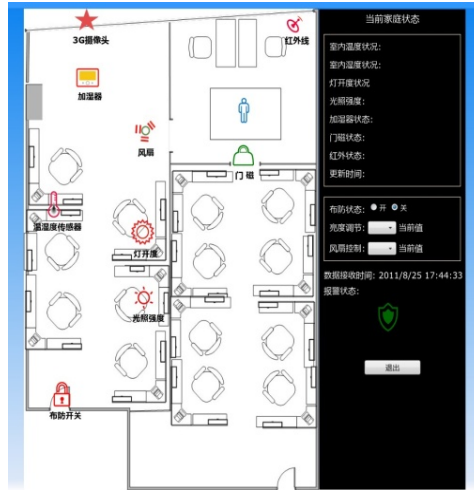
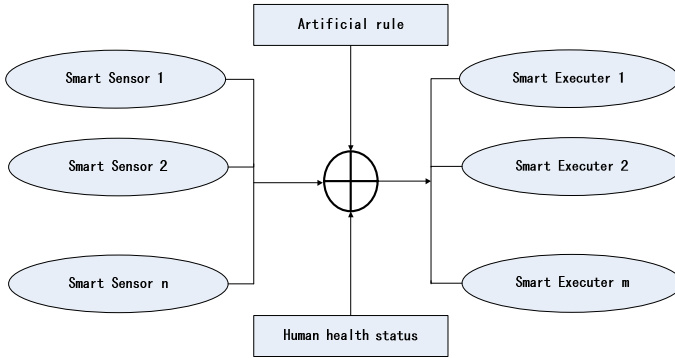


Fig. 5. Software running on mobile phone and web internet

## 5 Application Examples

Smart terminal is not to make people's lives lazier, but healthier and more comfortable. The Smart terminal is not super virtual entertainment system, but efficient energy management system. It is the home gateway that make them smarter . Home gateway help the in terms of linking their data together.

Smart terminals include smart appliances, smart home meters, and smart security equipment. Smart appliances mainly refer to intelligent home appliances such as smart air conditioner, smart refrigerator, smart microwave oven, smart ventilation system. Smart home meters mainly include water, electricity, and gas meter. Intelligent security device refers to fire safety, intrusion prevention system, such as anomaly detection, to provide a safe living environment. We propose a new structure of smart home. This structure contains sensing units, rule units and executer units. This defined classification makes smart home service more substantial. As it shown in Fig6. If there is a sensing unit, there should be something needed to be done. Those sensing unit direct smart executer to fulfill Specific mission under the guide of artificial rules and the status of human health. The artificial rules are the principles how make house environment more comfortable and healthier set by man. Three examples are given as follows.



**Fig. 6.** Mechanism of VCE system

**Smart Gate**

The Smart Gate’s sensing units are Magnetic device, Infrared device and cameras. The artificial rule is security settings inputted by human. And the executer is intrusion alarm system. When the security is set to high level, if one of the magnetic devices, infrared devices and cameras has detected intrusion target, alarm system will notify house owner in a way setted by the house owner previously.

**Smart air**

First, the smart air looks up the weather forecast for the user so that they can comfortably and adequately coordinate what they wear with the outside environment before they leave the house. Once the user returns home, the clothes that they have worn can be easily deposited to go through the simple laundry function. Second, it monitors the temperature and humidity and gas level of every room, and then before house owner come home, together with air conditioner and fanner, it can make the air comfortable. It can keep indoor relative humidity less than 65%; in summer wind speed of not less than 0.15 m / s, in winter not more than 0.3 m / sec.

**A. Smart health**

Smart health monitor vital signs of human, the result data is considered as strong rules that guide VCE running. The sensing data can also be stored for medical analysis, which can be a part of wireless hospital. This kind of wireless wearable health device monitors vital signs such as heart rate, blood pressure, blood oxygen, blood sugar, breathing, body temperature and physical activity status. The data collected by wireless wearable health device is uploaded to the home gateway and then to central data network services on the hospital’s server. Hospital’s server replies tips and advice on how to be healthier.

Also by sensing the body fingerprint of the user, smart health immediately recognizes which user is stepping on the smart health. It is easy to record who is at home and who is not. The display can give a message of choice to each user, whether it is ‘hello’ or an important message.



## B. Smart light

Smart light just feel the house light condition by illumination density detector, and control the lights and curtain motor.

## 6 Summary

On the basis of network, VCE forms an information interaction platform to make relationship between power consumption and comfortable Healthy life environment more reasonable, in a sustainable way. It is the first one that takes vital signs as the strong rule between sensing units and executers. Maybe it is a kind of wireless hospital. It integrates various technologies, products, services, from consumer appliances and equipment using advanced sensing, communications, and control technologies.

1. Low cost: sensor terminal nodes are all low cost device with zigbee chip as communication band. They are also low power consumer operating on GTS time slot.
2. Standardization: Various home components communicate with each other and standardized work has been done in the form of home gateway.
3. Cross platform: It can work with various platforms in the home system.
4. Self-organization: VCE can carry out complex configuration and management of the system by using 433M-10k fsk-modulation transmitter.
5. Expandability: The system can carry out software upgrade and functional expansion automatically.

## 7 Conclusion

In this paper, we propose a novel smart home system integrating many kinds of energy efficiency features. VCE has many distributed low-power sensors in the environment which is used for detection and better regulation of the human environment. These data are used for monitoring and analyzing consumed energy on device level in near real-time. VCE can adjust the temperature, humidity, light intensity volume and other information. And VCE can also detect human blood pressure through sophisticated sensors real-time status.

## References

1. Jahn, M., Jentsch, M., Prause, C.R., Pramudianto, F., Al-Akkad, A., Reiners, R.: The Energy Aware Smart Home. In: 2010 5th International Conference on Future Information Technology (FutureTech) (2010)
2. Zhao, X.: The Strategy of Smart Home Control System Design based on Wireless Network. In: 2010 2nd International Conference on Computer Engineering and Technology, ICCET (2010)

3. Papadopoulos, N., Meliones, A., Economou, D., Karras, I., Liverezas, I.: A Connected Home Platform and Development Framework for Smart Home Control Applications. In: 7th IEEE International Conference on Industrial Informatics, INDIN 2009 (2009)
4. NGN@Home ETSI initiative,  
<http://portal.etsi.org/at/ATNGNSummary.asp>
5. TEAHA project, <http://www.teaha.org>
6. EPERSPACE project, <http://www.ist-eperspace.org/>
7. AMIGO project, <http://www.hitech-projects.com/euprojects/amigo/>
8. SENSE project, <http://www.sense-ist.org/>
9. ATRACO project, <http://www.atraco.org>
10. HOME2015 programme, <http://home2015.i2r.a-star.edu.sg>

# Problem and Countermeasure on the Development of Mobile Electronic Commerce in China

Jiangjing Wang, Jiulei Jiang, and Feng He

School of Computer, Beifang Ethnic University, Yinchuan, Ningxia, China  
wj11@sohu.com, wj12010@gmail.com, hfeng@yahoo.cn

**Abstract.** Based on electronic commerce, mobile electronic commerce is a brand new method of business. It further transforms the means of exchange made by people in traditional business activities and has in-depth influence on the development of social economy. The article analyzes the current situation of the development of mobile electronic commerce in our country. It points out the problems faced by existing development of mobile electronic commerce in our country, including mobile payment, credit and law, and logistics, etc. The corresponding countermeasures are also put forward.

**Keywords:** Mobile electronic commerce, mobile payment, mobile phone identification policy.

## 1 Introduction

Mobile electronic commerce means the process and behavior to take advantage of various wireless intelligence terminals, such as mobile phone, palmtop, and other mobile communication facilities, to integrate with the internet and make electronic business activities [1]. Mobile electronic commerce is one of the trends of development of electronic commerce. The Mobile electronic commerce can customize according to the individualized demand and preference. The users can control the pattern to provide service and information. Compared with traditional electronic commerce, mobile electronic commerce can make the customers get the information of the whole network and the individualized service at any time and any place practically.

Since 2009, Chinese mobile internet entered the high-pace developing stage. According to the data of MIIT, in 2009, the amount of mobile phone terminal in China had achieved 0.747 billion [2]; according to the data of CNNIC, in the end of 2009, the amount of mobile phone netizen in China had surpassed 0.233 billion, while in 2010, the 3G users in China will surpass 85 million [3].

As the largest network retailer in Asia, Taobao launched the user platform wireless Taobao (m.taobao.com) in the wireless field, which was founded and operated in March, 2008. On March 30th, 2009, the “mobile electronic commerce industrial application project” supported by ten top industries was launched officially in Beijing. Among them, as the first project, Ufida mobile street application innovation obtained the popularization firstly. On June 27th, 2009, based on the mobile street (hapigo.cn), the advanced domestic mobile electronic commerce, “mobile electronic mall” was opened in Hangzhou. It symbolizes that the mobile electronic commerce in our

country begins getting into application period from definition period and platform period.

## **2 Main Problems Faced of Mobile Electronic Commerce**

### **2.1 Mobile Payment**

Mobile payment is a key factor that impacts the development of electronic commerce. There are four models of mobile payment as follows [4], operators leading, bank leading, third-party mobile payment platform company leading, and operators cooperate with bank or card organization to fund a company leading. The development and model of mobile payment and good development of industry chain are inseparable.

There is a problem of fusion between operators and banks for mobile payment commerce in China [5]. Currently the way of cooperation between operators and banks is that banking sector rents mobile network. The problem is that banks can't get much benefit from micro-payment so that they are unwilling to accept this case. However mobile payments are more micro-payments, such as buying beer, paying utility bills, and so on. Moreover to macro-payment, such as "cell phone felica", "cell phone banking", etc, the payment subject of the business is bank, so mobile operators just play the part to provide information channel to banks and users. The proceeds of mobile operators are only charge of notes or data traffic, while they should share the risk of user's transaction with banks during operation of mobile payment service, so that they are not initiative. In the use of "cell phone felica" for mobile payment, users have to pay for basic monthly service fee to not only mobile operators but also the bank. The transaction cost of "cell phone felica" users is much higher than cash transaction, as well as than credit card consumption, which becomes an important reason for restricting the development of mobile payment at present stage.

### **2.2 Credit and Law**

Credit problem is also primal problem faced with development of mobile electronic commerce. According to the survey, there are over 90% cell phone users received fraud notes in China, and more than half users query the safety of paying by cell phone. In a long past period, many illegal SP( telecom value-added service provider) trick cell phone users to customize expensive service through new business, for which users misgive about value added service, even less sending personal financial information by cell phone.

Related issues also exist in development of traditional electronic commerce. In 2009, online shopping dupery is ranked the first in complaints of internet service. Taobao, as one of largest C2C website in China, has occurred many fraudulent conducts since it was established. In China, mobile phone identification policy has not been put into practice well by now in 1st- and 2nd-grade city, even not any procedures in 3 rd- and 4th grade city. Anyone can purchase SIM card in news stall and ordinary stall without any identity authentication. In current stage of development in China, whether enterprise or personal, lack of credibility is ubiquity. Credit, as a

culture issue, needs to be cultivated for a long period, and an endless process that is really perfect and enjoys popular support.

In laws and regulations, Ministry of Information Industry, People's Bank of China, and China Banking Regulatory Commission are all Accelerating lawmaking about electronic commerce (including mobile electronic commerce). Some new laws and regulations are introduced, making the environments of mobile electronic commerce market normative. As it relates to themes departments, such as People's Bank of China, China Banking Regulatory Commission, mobile operators, CP, SP, as well as commercial bank, etc, which are links of electronic commerce, improvement of the law is a difficult and continuous process.

### **2.3 Logistics Facilities**

Logistics is the biggest bottleneck in the current development of mobile electronic commerce. Mobile electronic commerce can provide more favorable price by reducing intermediate link to lower the cost. If logistics are not handled well, saving cost may not offset the delivery expense. The logistics for intangible goods (such as software, music, online service) can be accomplished directly by internet. Problems are major in dispatching of great majority of tangible goods.

Internal logistics are major in three forms now: ① through the postal system. Mail delivery is a very common way in China, and the postal system has built a well-developed system, which can almost reach any inland corner, while the postal charge is generally above normal. ② Self-built dispatching system. Although it can dispatch goods to customers quickly, it needs huge investment, including storage facilities, transport vehicles, personnel, etc. Common enterprise can not reach it, and few China enterprises adopt this approach. ③ Draw support from third-party logistics enterprises. Since most online marketing enterprises don't have their own dispatching system in China, so dispatching enterprises arise like bamboo shoots after a spring rain, such as "Shentong express", "Yunda express", "Yuantong express", and "Shunfeng express", etc. Dispatching operators at present have realized the function of online reservation service and are perfecting the dispatching system ceaselessly in China. Online marketing enterprises and dispatching enterprises have formed a relationship of correlation dependence, and the greater online sales volume is, the huger dispatching market is. Time generally is within three to four days from ordering goods to receiving goods, and traffic expense is lower with carriage free increasing gradually. As professional logistics enterprises started to develop late in China and system is imperfect, the application of this way is restricted by geographical location. Also, service is imperfect, quality of deliveryman is lower, and customer satisfaction is poor.

## **3 Countermeasures That Should Be Adopted**

### **3.1 Government Leading and Unified Standard**

The solution of mobile payment depends on support and encouragement from government. Led by the government to develop uniform standards for mobile payments and introduce relevant encouraging policies to promote the rational

development of the industry chain, and to cultivate a favorable market environment. Second is market positioning, and developing micro-payment prior is conducive to the cultivation of mobile electronic commerce market. Taking the industry model of mobile payment in Korea as an example [6], SK, KTF, and LG, which are the three major mobile operators in Korea, are keeping a firm hand on the control of industrial chain of mobile payment, and occupying the dominant status of the industry chain. At present, three top mobile operators provide prepayment smart-card mobile phones with credit card and based on Felica standard. Among them, SK telecommunication, the Korean operator with good development of mobile payment, controls the entire industry chain. It cooperates with other links of the industry chain tightly and launches MONETA mobile payment brand; KTF also launches its own mobile payment brand K-merce; however, in 2003, LG telecommunication has launched the business of mobile phone bank. In addition, the successful fields of Korean mobile payment business mainly focus on mobile micro-payment. Such kind of success also refers to the uniform Korean mobile payment standard, accurate management of market positioning, and reasonable development of the industry chain.

Based on the current condition of our country, the restriction and kernel advantage of each link in the industry chain are different. The development pattern that is suitable for the domestic mobile phone payment field should take win-win combination of financial institutions and mobile operators as the major part, while the coordination and supporting of third-party payment service provider are the auxiliary. Take advantage of the comprehensive integration pattern, take full advantage of the resource of each layer, and finally achieve the win-win result. The Chinese mobile micro-payment market should be developed with priority. Further, unify mobile payment standard and relevant technique as well as business. Through issuing operating license, regulate the market. Thereby, further promote the healthy development of mobile electronic commerce.

### **3.2 Implement Mobile Phone Identification Policy**

Mobile phone identification policy is the most effective method to manage mobile phones. When purchasing the mobile phones, the customer has to register with ID card. Implement the registering mechanism of mobile phone number with one-user and one-network, and integration of mobile phone and phone number policy. If the mobile phone is missing, as long as reporting the loss to telecommunication department, the network will be stopped at once. The lost mobile cannot be used. Most importantly, implementing mobile phone identification policy can prevent purchasing the mobile phone with others' ID cards. At present, our country has established uniform ID information database around the whole country. As long as inputting the ID card number of the customer into the computer and connecting with information database, the identification of the ID card will be made at once. With the utilization of mobile phone identification policy, establish individual integrity record. It can restrict the sending of harmful messages effectively. If the criminal behavior is found, such as utilizing the mobile phone to defraud, the criminal suspect will be locked immediately and investigated.

The key of mobile phone payment is to establish perfect credit system, while mobile phone identification policy is the kernel link [7]. In order to popularize mobile

phone payment, the establishment of the credit account of mobile phone user has to be made, which do not only requires the achievement of technique, but also needs education to the awareness and idea of Chinese consumers. Some experts think that mobile phone identification policy can adopt certain interim measures. For instance, restrict some functions of the mobile phone service to the users who don't make identification registering. After the users complete identification registering procedure, the corresponding limitation will be canceled. In China, the network identification starts with games in order to protect the teenagers to grow up in healthy environment. However, in other aspects, there is still a long way to go, but what can be confirmed is that the identification policy of Chinese mobile electronic commerce in future is an inevitable trend. It needs to implement relevant corresponding laws and regulations aiming at ensuring the market order of mobile phone payment.

### **3.3 Accelerate the Development of Third-Party Logistics Industry**

In order to perfect the logistics of mobile electronic commerce in China, the key point is to complete the following two points.

First, pay attention to and accelerate the establishment of information system. In China, the development of modern logistics has certain difference with that of the developed countries and regions. The infrastructures and facilities of modern logistics in enterprises have certain scale. However, the idea, system, quality, variety, matching, and operating efficiency still need to improved greatly.

Second, take full advantage of the existing postal service network and platform. Cultivating the third-party logistics market in our country requires a long period. Nevertheless, utilizing existing postal service network and platform to develop logistics industry is a convenient passage in current stage. After dozens of years of development and construction, the Chinese postal service career has own certain foundation. In the aspect of article dispatching and network channel distribution, there's outstanding advantage. Specially, in the recent ten years, the informatization construction and the opening of postal saving business make China Post become the only service entity with the integration of information flow, capital flow, and material object flow. In logistic market, it is with obvious hardware advantage. Obviously, in service method and service quality aspect, China Post still needs to improve. In this way, it can achieve the service purpose with high efficiency and high quality required by modern logistics industry.

## **4 Conclusion**

With the large-scale commercial use of 3G network, it is predicted that the direct investment on 3G made by three top operators will surpass 300 billion yuan. The demand of enterprise and customers to mobile commerce application is also improved. It is an inevitable trend for modern business to move to 3G mobile internet. Considering the characteristics of mobile electronic commerce, mobile electronic commerce is very suitable for the popular application. Some experts predict that in the future period, mobile electronic commerce will become one of the main-stream developmental pattern of electronic commerce. At present, because of the restriction

of technique, matching, law and many factors in our country, mobile electronic commerce still requires a certain developmental period to be improved gradually.

**Acknowledgments.** This work is partially supported by NSF Grant #71061001 to F. He.

## References

1. Qin, C.-D., Wang, R.-L.: Mobile E-Commerce. People's Posts and Telecommunications Press (2009)
2. Chinese e-commerce research center. China Mobile e-commerce industry report (2009-2010), <http://b2b.toocle.com/zt/qg/>
3. Research, China Mobile e-commerce market research reports (2010), <http://www.ireserach.cn>
4. Varnali, K., Toker, A.: Mobile marketing research: The-state-of-the-art. *International Journal of Information Management* 30, 144–151 (2010)
5. ROTO. Mobile Internet User Experience (EB/OL) (September 9, 2010), <http://mobiforge.com/designing/blog/mobile-internet-user-experience>
6. Li, X.-F., Xiang, J.-Y.: South Korea's Development Research and Strategy Revelation of Mobile E-Commerce. *Tangdu Journal* 2, 118–120 (2004)
7. Mobile Web Best Practices Working Group. Mobile Web Best Practices 1.0 (EB/OL) (July 29, 2008), <http://www.w3.org/TR/mobile-bp/>



# Information Inquiry and Volume Analysis System on Urban Traffic Based on VB and MATLAB

Xiangcai Zhu, Yingkun Hou, Jian Xu, and Yuncai Luan

School of Information Science and Technology, Taishan University, Taian, China  
zhuxiangcai@126.com, sdkdhyk@163.com, xjmailbox@eyou.com,  
yos76@163.com

**Abstract.** In this paper, we carried out analysis, comparison, research on relevant technologies and methods, such as VB and MATLAB, and developed the information inquiry and volume analysis system. This system can realize: (1) inquiring the information of urban bus routes, stations, running schemes and the peculiar information of streets, (2) analyzing and forecasting traffic volume by Least Square Method (LSM) and Polynomial Fitting Method (PFM), and (3) qualitative analysis and quantitative analysis of the urban traffic faults.

**Keywords:** Urban traffic, volume, visual basic, matlab.

## 1 Introduction

The unobstructed factor affecting the urban traffic is much more diversified and complicated. Those problems such as the traffic jam, the vehicle manage, the safety of crossing are paid close attention by people. More scientific and modernized methods and technologies are used to manage the traffic[1]. Traffic information administration, urban 3-D simulation and area digitization have been developed gradually and improved urban traffic control level. [2]

We especially design the application module, including Traffic Information Inquiring Module (TIIM), Traffic Volume Analysis Module (TVAM) and Fault Tree Analysis Module (FTAM). We mainly realize three goals: (1) to realize inquiring about the information of urban bus routes, running schemes and the peculiar information of streets, (2) to realize analyzing traffic volume by Least Square Method (LSM), Polynomial Fitting Method (PFM), etc., and (3) to realize qualitative analysis and quantitative analysis of urban traffic faults. [3]

## 2 Relevant Technologies

MTALAB is the powerful soft that centers numerical value calculation, image administration and procedure development for an integral whole. The MATLAB 6.0 has important improvement on functions. It has debuted the SIMULINK module which is an integrated environment including model building, simulation and analyses, and it may improve the ability to learn through training for people about non-linearity and random system. It has developed exchanged module carrying out an

immediate data with the outside and opened up the road having got through MATLAB carrying out real time data analysis, data handling and hardware development. It has created high standard environment centering science calculation, image visualization and text handling.[4-6]

VB is the quick exploitation tool based on Windows, developed by the Microsoft Company in 1991. It is the soft system that has the certain advantage and potential in a lot of developing platforms, and its characteristic is visualization, event driving and human-computer interaction. The event driving is that order carries out not according to route but is decided by those events triggered from various running procedure. VB has good function of human-computer interaction. The main course of taking VB as the exploitation implement is that it has these characteristics. [7-8]

The system is developed by VB platform based on analyzing and studying the administration of current urban traffic and the development of relevant technologies. It can make the directors of urban traffic and passengers to provide traffic information in time, and realize scientific, intellectualized, modernized, and promote economic growth and accelerate urban progress. Applying the system has important practical significance.

### 3 System Structure and Functions

System structure contains hardware and software, and hardware includes computer, scanner, printer, sensor, etc. Software includes Windows XP, Office 2000, Matlab 7.0, VB 6.0, etc. System structure and function of every part shows as Fig. 1.

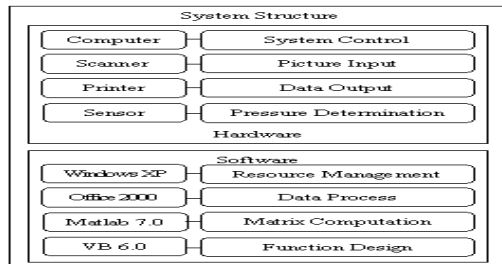


Fig. 1. System structure

VB is chosen to develop the system from the stability, reliability, pragmatism, tied in urban traffic real characteristic and developing advantage, and five modules are developed such as system setting (SS), file management (FM), data handling (DH), fun application (FA) and Auxiliary Function (AF), showed as Fig. 2. These modules have their own functions, and get in touch with and restrict, realize data sharing and exchanging. Among them the FA module includes traffic information inquiry (TII), traffic volume analysis (TVA) and Fault tree analysis (FTA).

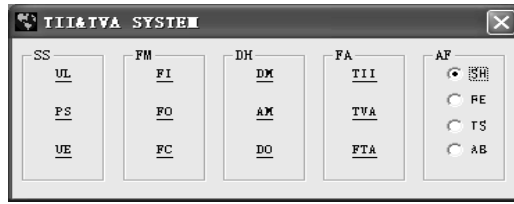


Fig. 2. System main interface

SS includes user-log (UL), parameter-set (PS) and user-exit (UE). When UL is hit, the password verification window will appear, including super user and domestic user. If password is input differently, the function is different after entering. PS includes route-set, background-set and database edition-set.

FM includes file-input (FI), file-output (FO) and file-change (FC). It can concretely realize entering, editing, browsing and revising the parameter and attribute of point, line or face. It can realize changing and storing different type of data. Data related to urban traffic include space data (e.g. location) and un-space data (e.g. unit character).

DH includes data-management (DM), attribute-management (AM) and data-output (DO). It can realize databases documents building-up, data inquiry, data administration, data back-up, and realize vector-grid change and data output.

FA can embody the synthesis and applicability. Its concrete function as follows: can realize information inquiry about all kinds of traffic data, including the routes inquiry, characteristic property inquiry, street inquiry, etc. The module is the very important part of the system. It can realize analyzing and forecasting traffic volume by Least Square Method, Polynomial Fitting Method, etc., and realize the forecast to the traffic accident by FTA method.

## 4 Developing Process

### 4.1 Information Inquiring Module

Information inquiring module is the key part of the system. When hitting the TII button on the system main interface, the TII interface will be appeared, showed as Fig. 3. It includes five parts such as rout-inquiring (RI), station-inquiring (SNI), scheme-inquiring (SEI), characteristic property-inquiring (CI) and street-inquiring (STI).

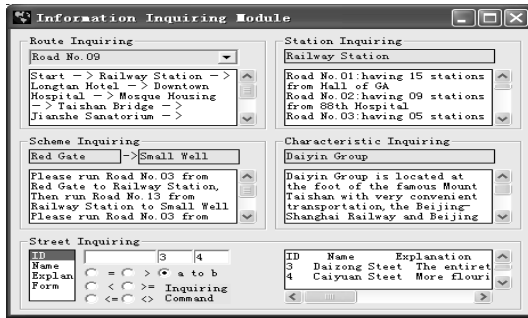


Fig. 3. Information inquiring interface

RI can realize choosing definite route. The combination frame contains all public bus routes. If Road No.09 was chosen, the text would show running route of Road No.09: Start→Railway Station→Longtan Hotel→Downtown Hospital→Mosque Housing→Taishan Bridge→Jianshe Sanatorium→Children's Paradise→Puzhao Temple→Training Center→Hushan Park→End, showed as the RI part on Fig. 3. The part developing codes as follows:

```

Dim SQLstring As String
Set rec = New ADODB.Recordset
SQLstring = "select * from biao where cc=" & Trim(Combo1.Text) & ""
rec.Open SQLstring, conn, 1, 3
If Not rec.EOF And Not rec.BOF Then
Text1.Text = "Start -> "
For jj = 1 To rec.Fields.Count 1
If Len(rec.Fields(jj)) > 1 Then
ZhanDian(jj) = rec.Fields(jj)
Text1 = Text1 & rec.Fields(jj) & "->"
Else:intMW = jj 1
Exit For:End If:Next jj:End If
.....

```

SNI can help users to inquire about the best place travelling by bus. When users import one station, the system will show all stations from this station. For example, if "Railway-station" was put in, there would appear the information below: Road No.01: Having 15 stations from Hall of GA; Road No.02: having 9 stations from 88th Hospital; Road No.03: having 5 stations from Tianwai Village; Road No.04: having 12 stations from TSU.

SEI can help users to choose the best scheme traveling by bus. For example, "Red Gate" was imported to the start-text and "Small Well" was imported to the end-text, there would appear two running schemes. One is that "please run Road No.03 from Red Gate to Railway Station, and then run Road No.13 from Railway Station to Small Well." The other is that "please run Road No.03 from Red Gate to Caiyuan, and then run Road No.13 from Caiyuan to Small Well."

CI can realize inquire about the peculiar information of peculiar station. When users inquire about the peculiar information of "Daiyin Group", there will show the information in the text: Daiyin Group is located at the foot of the famous Mount

Taishan with very convenient transportation, the Beijing-Shanghai Railway and Beijing-Fuzhou highway cross its one side.

STI can realize inquiring information about all streets and roads. Users can inquire the fields with ID, name or explanation. Users can also choose the concrete types such as ">", "=" and "a to b", etc. For example, if users want to inquire about information with ID, they may choose "a to b", and input 3 and 4 into the two texts, the correct information will show, as followed Fig. 3 right side. There will show all fields that meet to requirements to b from a.

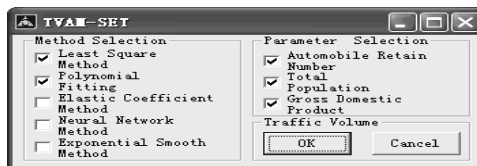
## 4.2 Traffic Volume Analysis Module

We develop the Traffic Volume Analysis Module, which mainly realizes model building, analyzing and forecasting to the element relation such as Traffic Volume (TV), Total Population (TP), Automobile Retain Number (ARN) and Gross Domestic Product (GDP), showed as Table 1. Assume that TP, ARN or GDP is independent variable, X, TV is dependent variable, Y, and we can set up the relation model between Y and X:  $Y = AX^3 + BX^2 + CX + D$ . Then we can select right algorithm to calculate the numerical value of A, B, C and D according to actual data, such as Least Square Method, Polynomial Fitting or other methods.

**Table 1.** The relation of all traffic elements

Year	ARN (104vehicle)	TP (105person)	GDP (1011rmb)	TV (108km/a)
1997	0.1	53.281	0.29130	29.56
1998	0.4	53.595	0.32214	29.92
1999	0.8	53.859	0.35631	31.42
2000	1.6	54.074	0.39640	32.10
2001	2.2	54.308	0.44800	33.64
2002	3.1	54.640	0.51520	36.16
2003	4.3	54.703	0.57190	38.68
2004	5.2	54.880	0.69194	41.73
2005	6.4	55.037	0.85566	42.19
2006	7.8	55.170	1.01220	42.27

We choose LSM and PFM to calculate and analyze to traffic volume. When hitting the button TVA, the TVAM-SET interface will be appeared, showed as Fig. 4, which includes method selection, parameter selection and function buttons.



**Fig. 4.** TVAM-SET interfaces

The result of handling traffic data with MATLAB will be sent back to TVAM-RUN interface for browse and analysis, showed as Fig. 5. Fig. 5a shows cooperation relation between VF and GDP, Fig. 5b shows cooperation relation between VF and TP, and Fig. 5c shows cooperation relation between VF and ARN. Obviously, Fig. 5a and Fig. 5c show very good to fit result, and Fig. 5b shows not ideal to fit result by LSM. From many times experiment and changing the function "F = xx (1) \* xdata. \* xdata. \* xdata + xx (2) \* xdata. \* xdata + xx (3) \* xdata + xx (4)" to the function "F = xx (1) \* xdata. \* Cos (xdata) + xx (2)", the result is very ideal, as shows Fig. 5d.

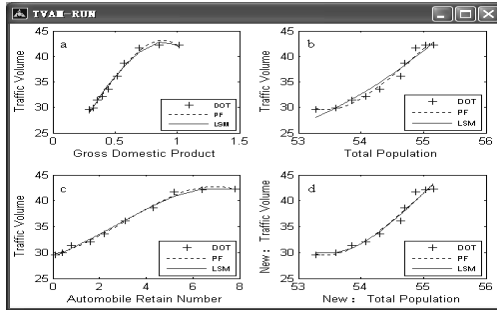


Fig. 5. TVAM-RUN interfaces

The part MATLAB commands called by the system as follows:

```

%Functions%
function F = myfun(xx,xdata)
F = xx(1)*xdata.*xdata.*xdata+xx(2)*xdata.*
xdata+xx(3)*xdata+xx(4);end;
function F = myfun(xx,xdata)
F =xx(1)*xdata.*cos(xdata)+xx(2);end
%Gross Domestic Product and Traffic Volume
xdata=[0.29130 0.32214 0.35631 0.39640 0.44800 0.51520 0.57190 0.69194
0.85566 1.01220];ydata=[29.56 29.92 31.42 32.10 33.64 36.16 38.68 41.73 42.19
42.27];
pcoef=polyfit(xdata,ydata,3);
xx=0.2913:0.00001:1.0122;yy=polyval(pcoef,xx);subplot(2,2,1);plot(xdata,ydata,'b
+');hold on;plot(xx,yy,':k');xx = lsqcurvefit(@myfun,[0.5 0.5 0.5 0.5],xdata,
ydata);hold on;
plot(xdata,xx(1)*xdata.*xdata.*xdata+xx(2)*xdata.*xdata+xx(3)*xdata+xx(4),'r');
legend('\fontsize{6}DOT',\fontsize{6}PF',\fontsize{6}LSM',4);xlabel('Gross
Domestic Product');ylabel('Traffic Volume')
%Total Population and Traffic Volume%
xdata=[53.281 53.595 53.859 54.074 54.308 54.640 54.703 54.880 55.037 55.170];
ydata=[29.56 29.92 31.42 32.10 33.64 36.16 38.68 41.73 42.19 42.27];
pcoef=polyfit(xdata,ydata,3);xx=53.281:0.001:55.170;yy=polyval(pcoef,xx);subp
lot(2,2,2);
.....

```

The data handling part comes true by transferring MATLAB commands and VB mainly realizes interface programming, data calling, method selecting, etc.

This module mainly curves fitting to the scattering data between every factor and traffic volume in order to grasp the regularity and predictability and provides the theory basis and the science method for the communications department in the field of traffic volume forecast, traffic control, communications construction, urban planning, etc.

### 4.3 Fault Tree Analysis Module

If user hit "FTA" menu of main surface, there will appear FTA interface, showed as Fig. 6, including six parts that are qualitative analysis, quantitative analysis, method selection, cause classification, element & weight and safety evaluation.

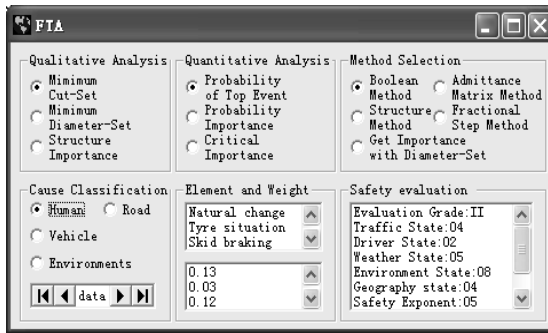


Fig. 6. FTA interfaces

Qualitative analysis includes minimum cut-set calculating, minimum diameter-set calculating and structure importance calculating. Quantitative analysis includes top probability of top event calculating, probability importance calculating and critical importance calculating. Method selection includes Boolean method, admittance matrix method, structure method, fractional step method and to get importance with diameter-set. Cause classification human cause, road cause, vehicle cause and environment. It also can realize browsing data through the data-control. Element & weight includes various element listing and corresponding weight. Safety evaluation embodies the corresponding valuation grade and the accurate valuation conclusion. For example, after cause classification and element & weight are ascertained, the window of safety evaluation will appear the following information: Evaluation Grade: II, Traffic State: 04, Driver State: 02, Weather State: 05, Environment State: 08, Geography state: 04, Safety Exponent: 05. The application of this module will reduce occurrence probability of urban traffic accident, and improve scientific and accurate degree of analyzing and forecasting management level of promoting traffic.

## 5 Conclusion

The system mainly selects suitable platform to realize station locate, information inquire, scheme choose, looks for good method to curve fitting traffic volume, and uses FTA to realize qualitative analysis and quantitative analysis to the urban traffic faults. The application will produce fine social and economic benefit, economize large amount of manpower and material resources, and improve traffic running efficiency and managing level.

The system has the following characteristics: having the fine ability to manage and search traffic information; having the ability of dynamically displaying the public bus routes; having the good function of curving fitting, and so on. The application and upgrade of the system will realize making use of the information effectively, improving the visualization degree of different information, managing urban traffic high-effectly, and promoting urban economic development and improving the living standard of people.

**Acknowledgments.** This research has been partially supported by project No.J05G53 of Education Department of Shandong Province. We like to express our appreciation for the valuable suggestions from the referee and the editor of this journal which significantly improved the quality of the presented paper.

## References

1. Michelle, C., David, S., Nikolay, T.: Urban traffic systems modelling methodology. *International Journal of Production Economics* 99(1/2), 156–176 (2006)
2. Florence, B., Sophie, M., Jean-Claude, P.: The real-time urban traffic control system CRONOS: Algorithm and experiments. *Transportation Research, Part C, Emerging Technologies* 14C(1), 18–38 (2006)
3. Masakazu, Y., Kazunori, M., Yukio, F., et al.: An agent-based approach to urban traffic flow simulation. The Institute of Electronics. Technical Report of IEICE Information and Communication Engineers 105 (501), 267–272 (2005)
4. Li, M.Z.: The road traffic analysis based on an urban traffic model of the circular working field. *Acta Mathematicae Applicatae Sinica* 20(1), 77–84 (2004)
5. Sigurour, F.H., Roland, C., Andreas, P., et al.: A high-resolution cellular automata traffic simulation model with application in a freeway traffic information system. *Computer-Aided Civil and Infrastructure Engineering* 19(5), 338–350 (2004)
6. Mihaela, P., Alexandru, B., Mircea, D.: Matlab GUI application in energetic performances analysis of induction motor driving systems. *WSEAS Transactions on Advances in Engineering Education* 3(5), 304–311 (2006)
7. Yang, Y.J., Deng, H.Y., Li, X.: Simulation of screening process based on MATLAB/Simulink. *Journal of China University of Mining and Technology* 16(3), 330–332, 337 (2006)
8. Lin, X.Y., Liu, A.D., Li, Y., et al.: A MATLAB programming for simulation of X-ray capillaries. *Applied Mathematics and Computation* 172(1), 188–197 (2006)



# Multivariate Analysis on System Reliability and Security

Xiaoyan Wu<sup>1</sup>, Zihui Che<sup>2</sup>, and Jing Jia<sup>3</sup>

<sup>1</sup> Department of Foreign Language Teaching and Research, Hebei University

<sup>2</sup> Department of Mathematics and Computer, Baoding University

<sup>3</sup> Baoding Electric Power Corporation, State Grid

**Abstract.** In order to ensure the reliability and security of electric power system, the faults must be removed exactly and quickly. The fault location method is one of key issues for the researches of electric power system. Extracting fault feature information from large amount of synchronized data will be a new challenge for locating of faulted components or faulted area. Multivariate analysis theory is superior in reducing dimensions and simplifying data structure featured without key information loss. In this paper, according to multivariate analysis theory, the reliability and security of complex power system has been accomplished.

**Keywords:** System reliability, Security, Multivariate analysis, WAMS, PMU.

## 1 Introduction

Electric power system is one of the most complex artificial systems in this world and its safe, steady, economical and reliable operating plays a very important part in guaranteeing socioeconomic development, even in safeguarding social stability. With the gradually deregulation of power system and the widely introduction of competitive mechanism, the current power system devices are operated close to their physical limits day by day, and the power transmission grids are interconnected more and more tightly. In the circumstances, the relay protection, which always plays the role as the first defending line, will undertake more responsibility for security and stability of power system. Up to now, the principles of main protection have a well development in theoretical research and field inspection. However, the backup protection, as an efficient supplement for primary one, is still puzzled by a series of problems, such as complicated in cooperation relationship, long in operation time, difficult in setting and configuration, noneffective in identification of flow transferring and so on [1-3].

Therefore, in many worldwide blackouts the traditional backup protections have acted “correctly” according to the existing design philosophy of protection, but which actually added fuel to the fire in accelerating breakdown of region power system. It is the root cause that only the local information of power grid is used in the conventional backup protection, which is failed in adopting more valid status data from a more wide area in design and implementation.

Currently, the new measure system as WAMS/PMU (WAMS, Wide Area Measurement System; PMU, Phasor Measurement Unit) could be used to obtain the global operation condition information of power system [4,5]. In other words, the

synchronized phasor data of nodal voltage and branch current with the same time stamp is becoming one new data source for power system analysis and design. In this way, the novel backup protection based on wide area synchronized information has attracted more attentions in field of power system relay protection in the past two decades. The basic rule of this novel protection is to locate the faulted components as soon as possible within the inherent time delay of backup protection, and in this locating process the synchronized phasor data uploaded by the widely installed PMUs is used for its short time refresh period. According to the fault location results, the corresponding tripping strategy could be schemed. Once the primary protection fails in acting, the backup one will operate to isolate the related faulted components by tripping the schemed circuit breakers set.

Hence, the fault location method is the one of key issues for novel wide area backup protection. Extracting fault feature information from large amount of synchronized data will be a new challenge for locating of faulted components or faulted area.

Multivariate analysis theory is superior in reducing dimensions and simplifying data structure featured without key information loss. In order to code with the huge quantity of synchronized information supplied by WAMS, the multivariate analysis theory could be used as a powerful mathematical tool. As to fault feature study in power system, reference papers[6-10] have provided a new way for introducing cluster analysis, pattern recognition and principal component analysis into the related fields of power system, especially for fault location needed in novel wide area backup protection, and the results of above methods are encouraged. In this paper, the multivariate analysis on the reliability and security in electric power system will be discussed carefully.

## 2 Nomenclature

$G_1, G_2$	Populations
$\Sigma$	Covariance matrix
$f_i(x)$	Probability density function
$W_i(x)$	Classification function
$\mu_i$	Mean vector
$\overline{x^{(i)}}$	Training sample
$p_i$	Prior probability

## 3 Multivariate Analysis Theory

Discrimination and classification are multivariate techniques concerned with separating distinct sets of objects and with allocating new objects to previously defined groups. Discriminant analysis is rather exploratory in nature.[11] In Bayesian discriminant analysis, Suppose  $G_1, G_2$  are two different  $p$ -dimensional populations which are known, the covariance matrix of these populations is  $\Sigma$ , and their probability density functions are  $f_1(x)$  and  $f_2(x)$  respectively: [17]

$$\begin{cases} f_1(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)} \\ f_2(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2)} \end{cases}$$

Let us differentiate them,

$$\begin{cases} \ln f_1(x) = -\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) \\ \ln f_2(x) = -\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x - \mu_2)^T \Sigma^{-1} (x - \mu_2) \end{cases}$$

According to the Bayesian optimal dividing rules, one can obtain,

$$\begin{aligned} R_1 &= \{x : p_2 f_2(x) \leq p_1 f_1(x)\} \\ &= \left\{x : -\frac{1}{2} (x - \mu_2)^T \Sigma^{-1} (x - \mu_2) \right. \\ &\quad \left. + \ln p_2 \leq -\frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) + \ln p_1 \right\} \end{aligned}$$

wherein  $p_1$  and  $p_2$  are the prior probability of  $G_1, G_2$  appearance. And the classification functions can be expressed as,

$$\begin{cases} W_1(x) = a_1^T x + b_1 \\ W_2(x) = a_2^T x + b_2 \end{cases}$$

So,

$$\begin{aligned} R_1 &= \{x : W_1(x) \geq W_2(x)\} \\ R_2 &= \{x : W_1(x) < W_2(x)\} \end{aligned}$$

In this place, the Bayesian discriminant criterion of two populations can also be expressed as,

$$\begin{cases} x \in G_1, & \text{If } W_1(x) \geq W_2(x) \\ x \in G_2, & \text{If } W_1(x) < W_2(x) \end{cases}$$

If  $\mu_1, \mu_2$  and  $\Sigma$  are unknown, then we can use the training samples of  $G_1, G_2$  to estimate them. In fact, the estimates of  $\mu_1, \mu_2$  are just the training samples' mean vectors  $\overline{x^{(1)}}, \overline{x^{(2)}}$ , and the estimates of  $\Sigma$  are just the training samples' covariance matrix  $S$ , then the classification functions are:

$$\begin{cases} \hat{W}_1(x) = \hat{a}_1^T x + \hat{b}_1 \\ \hat{W}_2(x) = \hat{a}_2^T x + \hat{b}_2 \end{cases}$$

Furthermore, the posterior probability can be expressed as,

$$\begin{cases} P(G_1 | x) = \frac{p_1 f_1(x)}{p_1 f_1(x) + p_2 f_2(x)} \\ P(G_2 | x) = \frac{p_2 f_2(x)}{p_1 f_1(x) + p_2 f_2(x)} \end{cases}$$

and

$$\begin{cases} p_1 f_1(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}d_1^2(x)} \\ p_2 f_2(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}d_2^2(x)} \end{cases}$$

wherein

$$\begin{cases} d_1^2(x) = (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) - 2 \ln p_1 \\ d_2^2(x) = (x - \mu_2)^T \Sigma^{-1} (x - \mu_2) - 2 \ln p_2 \end{cases}$$

So, the estimates of posterior probability:

$$\begin{cases} \hat{P}(G_1 | x) = \frac{e^{-\frac{1}{2}\hat{d}_1^2(x)}}{e^{-\frac{1}{2}\hat{d}_1^2(x)} + e^{-\frac{1}{2}\hat{d}_2^2(x)}} \\ \hat{P}(G_2 | x) = \frac{e^{-\frac{1}{2}\hat{d}_2^2(x)}}{e^{-\frac{1}{2}\hat{d}_1^2(x)} + e^{-\frac{1}{2}\hat{d}_2^2(x)}} \end{cases}$$

And the ultimate Bayesian discriminant criterion can be expressed as,

$$\begin{cases} x \in G_1, & \text{If } \hat{P}(G_1 | x) \geq \hat{P}(G_2 | x) \\ x \in G_2, & \text{If } \hat{P}(G_1 | x) < \hat{P}(G_2 | x) \end{cases}$$

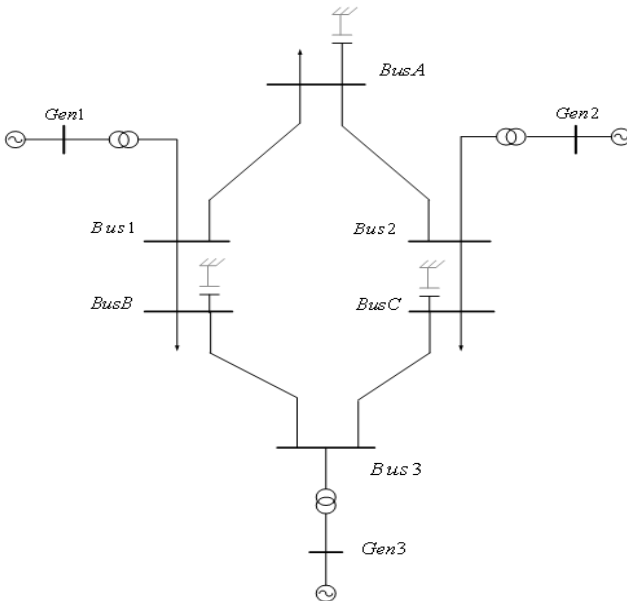
### 4 Multivariate Analysis on System Reliability and Security

In many cases, we have certain prior information about the distribution of research objects. Bayesian discriminant analysis is just utilizing this kind of prior information.

The basic idea is that all of  $k$  categories can be regarded as mutually exclusive subdomains in space, and each observation is a point in this space. Under the precondition of considering prior probability, one can use Bayesian discriminant analysis to construct discriminant functions based on some rules, and calculate the probability of some sample falls into each subdomain respectively. The category with maximum probability is considered to be the class that the sample belongs to the test examples used in the present research, such as IEEE nine nodes with three generators is usually employed and studied in the field of power system and its automation, which is also internationally accepted as the standard power system test cases.

Figure 1 presents the electric diagram of IEEE nine nodes system. In the structure of electricity grid, Bus1 appears single-phase to ground fault. For IEEE nine nodes system, the node negative sequence voltages at  $T_{-1}, T_0$  (Fault) and  $T_1$  three times can be obtained. According to Bayesian discriminant analysis principles, one calculates the classification function coefficients. Finally the results of posterior probability and discriminant classification based on Bayesian discriminant analysis have been obtained, see Table 1.

Similarly, suppose Bus1 appears double-phase to ground fault. According to the node negative sequence voltages, the results of posterior probability and discriminant classification based on Bayesian discriminant analysis have also been listed in Table 2.



**Fig. 1.** IEEE nine nodes system

**Table 1.** The posterior probability and classification of single-phase fault

Node	Classification	Posterior probability (Fault population)	Posterior probability (Normal population)	Discriminant classification
<i>Bus1</i>	F	0.99096	0.00904	F
<i>Bus2</i>	N	0.00275	0.99725	N
<i>Bus3</i>	N	0.00451	0.99549	N
<i>BusA</i>	N	0.41227	0.58773	N
<i>BusB</i>	N	0.55977	0.44023	F
<i>BusC</i>	N	0.00239	0.99761	N
<i>Gen1</i>	N	0.00931	0.99069	N
<i>Gen2</i>	N	0.00029	0.99971	N
<i>Gen3</i>	N	0.00066	0.99934	N

**Table 2.** The posterior probability and classification of double-phase fault

Node	Classification	Posterior probability (Fault population)	Posterior probability (Normal population)	Discriminant classification
<i>Bus1</i>	F	0.99095	0.00905	F
<i>Bus2</i>	N	0.00274	0.99726	N
<i>Bus3</i>	N	0.00453	0.99547	N
<i>BusA</i>	N	0.41134	0.58866	N
<i>BusB</i>	N	0.5605	0.4395	F
<i>BusC</i>	N	0.0024	0.9976	N
<i>Gen1</i>	N	0.00935	0.99065	N
<i>Gen2</i>	N	0.00029	0.99971	N
<i>Gen3</i>	N	0.00066	0.99934	N

To sum up the above Bayesian discriminant classification results, the misjudgment ratio is,

$$\hat{p}_r^* = \frac{1}{9} = 0.111.$$

Consequently, for single-phase to ground fault and double-phase to ground fault in IEEE nine nodes system, Bayesian discriminant analysis has confirmed fault position. Therefore, based on multivariate analysis theory, the reliability and security of electric power system has been assured.

### Conclusions

The fault in complex electric power system is inevitable. In many worldwide blackouts the traditional backup protections have acted “correctly” according to the existing design philosophy of protection, but which actually added fuel to the fire in accelerating breakdown of region power system. The fault location method is one of

key issues for novel wide area backup protection. Extracting fault feature information from large amount of synchronized data will be a new challenge for locating of faulted components or faulted area. In this paper, based on multivariate analysis theory, the reliability and security in electric power system has been assured.

**Acknowledgments.** This research was supported partly by the Natural Science Foundation of Hebei Province.

## References

1. Horowitz, S.-H., Phadke, A.-G.: Third zone revisited. *IEEE Transactions on Power Delivery* 21, 23–29 (2006)
2. Phadke, A.-G., Thorp, J.S.: *Synchronized Phasor Measurements and Their Applications*. Springer (2008)
3. Phadke, A.-G., Thorp, J.-S.: *Computer Relaying for Power System*, 2nd edn. John Wiley & Sons Ltd., Chichester (2009)
4. Wang, C., Dou, C.-X., Li, X.-B., Jia, Q.-Q.: A WAMS/PMU-based fault location technique. *Electric Power Systems Research* 77, 936–945 (2007)
5. Peng, J.-N., Sun, Y.-Z., Wang, H.-F.: Optimal PMU placement for full network observability using Tabu search algorithm. *International Journal of Electrical Power & Energy Systems* 28, 223–231 (2006)
6. Wang, Z.-P., Zhang, Y.-G., Zhang, J.-F.: Recent research progress in fault analysis of complex electric power systems. *Advances in Electrical and Computer Engineering* 10, 28–33 (2010)
7. Zhang, Y.-G., Wang, Z.-P., Zhang, J.-F.: Fault discrimination using synchronized sequence measurements under strong white gaussian noise background. *International Journal of Emerging Electric Power Systems* 12, 1–15 (2011)
8. Zhang, Y.-G., Wang, Z.-P., Zhang, J.-F., Ma, J.: Fault localization in electrical power systems: A pattern recognition approach. *International Journal of Electric Power & Energy Systems* 33, 791–798 (2011)
9. Zhang, Y.-G., Wang, Z.-P., Zhang, J.-F., Ma, J.: PCA fault feature extraction in complex electric power systems. *Advances in Electrical and Computer Engineering* 10, 102–107 (2010)
10. Zhang, Y.-G., Wang, Z.-P., Zhang, J.-F.: Universal characteristics of faults in complex power system. *International Review of Electrical Engineering* 6, 1474–1482 (2011)
11. Mei, C., Fan, J.: *Data Analysis Methods*. Higher Education Press, Beijing (2006)

# Design and Application of a Network Platform for Electronic Contract Making

Yang Jianzheng and Xia Yunchao

Business School, University of Shanghai for Science and Technology,  
No. 516, Jungong Road, Shanghai, China  
cnyangjz@163.com, supercosmo@qq.com

**Abstract.** Accompanying with the popularity of on-line business, e-business development in China has to break the bottleneck of traditional off-line mode. Therefore, seeking for a more economic, effective and efficient manner with security in contract making seems to be inevitable. The paper explains the status of current business negotiation, analyzes the difficulties in the popularizing of applying the electronic signature into commercial activities, and then raises the pioneer Network Platform for Electronic Contract Making as well as its applications. The study concentrates on the following highlights of the platform: firstly, the innovative application of electronic signature on contract making by integrating the digital signature technology into online negotiation and chatting system; secondly, realizing conservation of electronic contracts in an independent third party for future contract disputes; lastly, developing and forming a set of standard electronic contract templates in tree structures for the first time.

**Keywords:** e-business, electronic contract, electronic signature, negotiation, innovative application.

## 1 Introduction

This section mainly deals with the background of the research, including the existing problems, status of current electronic contract negotiation as well as breakthrough point of the application of electronic signature.

### Existing Problems

Negotiation, such as commercial contract negotiation and international dispute negotiation, is a common issue existing in economic and social fields, being one of the most important methods to resolve controversies and conflicts. In terms of business activities, negotiation is a common manner to reach an agreement among willing parties. According to a research from Foroughi, in organizations, negotiation activity occupies a manager about 20% of his total working time [1]. Moreover, negotiators can rarely reach a satisfied agreement with all related parties. All above deficiencies waste not only social resources, but also productivities and opportunities for innovation. Therefore, the issue of improving negotiation efficiency is raised.



On the other hand, in recent years, with the rapid development of e-business, globalization of e-commerce emerges inevitably. Additionally, modern companies trend to be scalization, conglomeration, transregionalizaion as well as internationalization, which all contribute to an acute exploitation of a better negotiation manner with characteristics of economy, efficiency, effectiveness and security.

### **Status of Current Electronic Contract Negotiation**

Many areas of commercial activities have been conducted electronically, however, as a key process of e-business transaction, electronic contract making still progresses slowly. Reasons may be defined as follows:

Firstly, existing electronic contracts are in fixed formats, without a proper manner for interactive negotiation among contractual parties. For example, when you download software from a website, you will probably be asked to tick a box and confirm a fixed format statement as a disclaimer for infringement of copyrights. The disclaimer is a kind of electronic contract called click contract.

Secondly, no mature electronic contract negotiation and signing software is available nowadays. Although text, audio and video exchanging technologies have been sophisticated and applied in online chatting software, such as MSN, ICQ and Skype, nothing to do with existing electronic contracts.

Thirdly, in current, traditional standard contracts in total number of 2400 are expected to be transferred into electronic ones, leaving great amounts work to do. However, no electronic contract templates have been developed to meet the potential demands.

### **Breakthrough Point of the Electronic Signature Application**

In 5 July 2001, "UNCITRAL Model Law on Electronic Signatures with Guide to Enactment", being drafted by United Nations Commission on International Trade Law (UNCITRAL) was adopted. Meanwhile, in 28 April 2004, "Electronic Signature Law of the People's Republic of China" was approved by the National People's Congress. However, there has been no breakthrough in the application of electronic signature so far, resulting in tens of electronic certification institutes being at a loss in China. Although electronic transactions have been commonly applied in stock exchanges, no better application points have been explored in other fields. Therefore, electronic signature is still distant to our people and the reasons may be concluded as follows:

First of all, lack of propaganda. People even have no impression on the function of electronic signature; second, no breakthrough point is selected in its application; last but not least, technology of electronic signature has not been well developed.

All in all, in terms of current situation, the breakthrough point of electronic signature application should be the procedures of business negotiation and contract signing where electronic signature can be conducted frequently and exert its utility to maximum. The following study raises the pioneer Network Platform for Electronic Contract Negotiation and Signing Service, and attempts to apply the 2-dimensional bar code associating with the platform to a commercial environment.

## 2 Highlights of the Platform

The platform provides users a whole new concept of business negotiation through the creative application of electronic signature on contract making, realizing the equal legal effect between paper contracts and electronic ones.

### **Seamless Integration between Electronic Signature and Electronic Contract**

The system applies the 2-dimensional bar code encryption technology from Shanghai Electronic Certification Authority Center (CA). Through customizing Java classes, the system successfully realizes the seamless integration between electronic signature and electronic contract and ensures that all electronic contracts are in characteristics of privacy, completeness and non-repudiation.

### **Secured Conservation of Electronic Evidence**

Our platform invites Shanghai Trade Point which is affiliated to Shanghai Municipal Commission of Commerce as an independent third party to conserve the electronic contracts. Associating with the technologies such as VPN data transmission, timestamp, network-attached storage (NAS), storage area network (SAN), etc. the platform resolves the problems including contract loss, unauthorized modification and disclosure on contracts. Moreover, it provides strong evidence in subsequent contract disputes between contractual parties.

### **Standard Electronic Contract Templates**

It seems that few electronic contract template is widely used in domestic or abroad in business negotiations. The platform develops a set of electronic contracts, covering almost all types of common contracts. What's more? The system provides absolute flexibility to users. They may either benefit from these existing contract templates in the system or enjoy editing and forming their own templates through paste the text from Microsoft word or excel, where pictures are also acceptable.

## 3 Design of the Platform

As have introduced the background and highlights of the Network Platform for Electronic Contract Making, this section mainly deals with the detailed design of it, including the design of system flow, introduction of function modules, clients' function display as well as the B/S structure of system.

### **System Flow**

The Network Platform for Electronic Contract Making aims to assist contractual parties to reach an agreement by means of an electronic contract. Therefore, the system flow strengthens the order of following aspects, including contractual party relation establishment, contract term negotiation, document generation, electronic signature application as well as contract saving and printing. Figure 1 gives the detailed flow of the system.

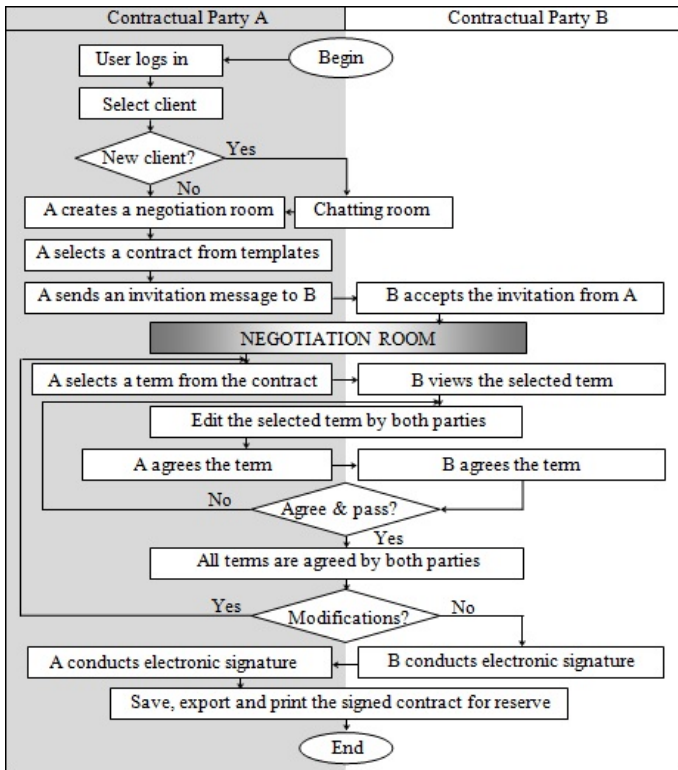


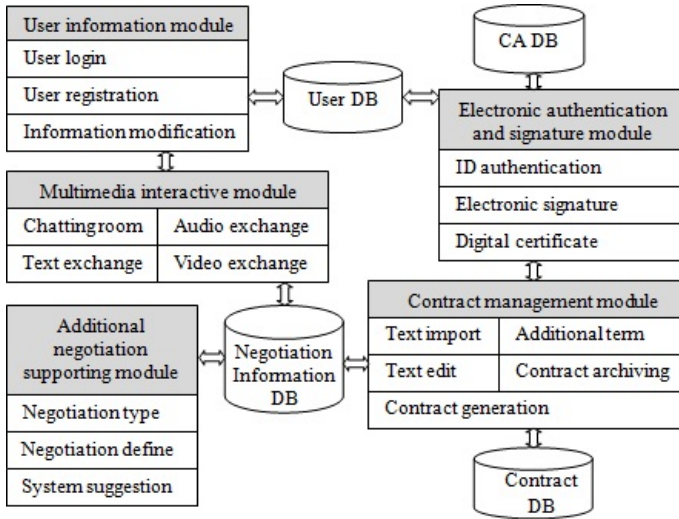
Fig. 1. The flow chart indicates the main procedures of electronic contract making in terms of user A. The terms of the contract should be negotiated under a mode of “term by term” according to the flow. Any modifications on the contract by A will be rejected by the system automatically and be invalid after B conducting his electronic signature, and vice versa.

**System Module Introduction**

According to the above flow of contract making, together with the overall function of the Network Platform for Electronic Contract Making, the system is abstracted into five main function modules, including the multimedia interactive module, user information module, contract management module, electronic authentication and signature module, as well as the additional negotiation supporting module (see Fig.2).

The basic ideas of the design of the function modules are concluded as follows:

- Realize text, audio and video exchange for both contractual parties; and
- Provide optional electronic contract templates; and
- Realize searching, browsing and editing functions of electronic contract templates; and
- Realize electronic signature application, electronic contract saving, archiving and security authentication function; and
- Endow electronic contract with the equal legal effect against traditional paper contract.



**Fig. 2.** The figure shows the structure and relation of all function modules of the system, together with four main DBs

**Function Display**

In terms of client’s screen display of above functions, the window of the core negotiation room will be divided into four parts:

*Electronic Contract Template Area*

Only titles of all terms and conditions of the contract will be presented in a tree structure in this area. A tree structure is where titles of contract terms are listed as “branches” of a “tree” at different levels and grades, acting as a guide for users to import content of terms into Contract Term Editing Area by clicking them. In terms of customized contract templates, the tree structure may also be customized and the system provides absolute flexibility to users. That is to say, which terms or conditions will be the “branches” of the tree ups to users themselves.

*Multimedia Interactive Area*

Users should chat with each other to exchange their views during the negotiation. Multimedia Interactive Area meets the demands of users by providing text, audio and video exchange functions.

*Contract Term editing Area*

Users may select the titles from the tree structure and edit the content of corresponding terms here. The system distinguishes the chatting area from the term editing area, as the final document may only be composed of contract terms.

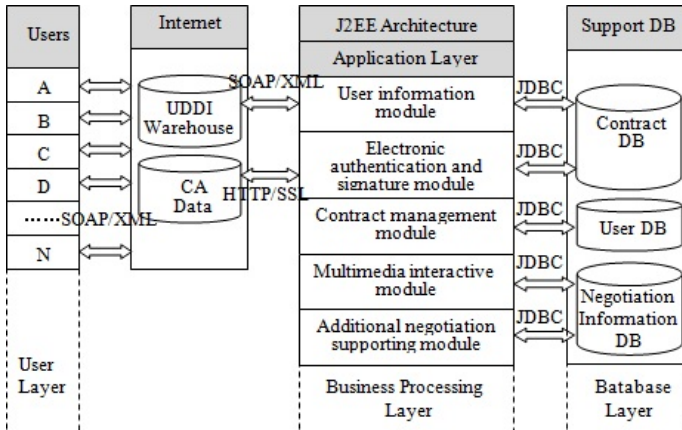
*Contract Previewing Area*

Those edited terms will be shown in the area as a preview of the final document. Any modifications of terms presented here will still be welcomed. Users may either click them directly or select the title of them at Electronic Contract Template Area to return them back to Contract Term editing Area for further consideration.

The final document can either become a legal binding contract immediately by clicking the electronic signature button and typing encryption code, or leaving an unfinished document to be off-line hand-signed.

**B/S Structure of the Platform**

The platform applies a 3-layer B/S structure based on the J2EE framework (see Fig.3).



**Fig. 3.** The figure indicates the 3-layer B/S structure of the platform

User layer includes system’s logic of display, whose task is to send the service request from the client to the web server on Internet. The web server may return the information back to the client after verifying his identification via a predetermined protocol (i.e. SOAP in our system), and then, the client will receive the information in XML and display it.

Business processing layer is provided with application extensions, including the system’s logic of transaction processing and its main task is to receive clients’ requests. Information will be transferred from database server to the client after connecting the application extensions and databases through SQL.

Database layer includes the system’s logic of data processing. Its main task is to receive the request from web server, realize database inquiry, modification and updating, and return the result back to web server through JDBC protocol.

**4 Operation Prospects**

The platform aims to be applied in B2B websites as a third party platform because only commercial entities demand for trade contracts. Those individual customers using B2C or C2C websites may only ask for invoices rather than formal contracts. Additionally, some other issues considering platform’s future prospects are raised: First, acting as a new rising platform confidentiality of the electronic contracts maybe doubted; Second, the security and authority of the electronic signature maybe questioned by users; Third, lack of propaganda and potential users. To resolve above problems, our platform is based on a joint operation mode:

### **E-commerce Research Center of Shanghai University for Science and Technology**

As the main R&D entity, the center raises the basic idea and realizes the establishment of the platform.

### **Shanghai Trade Point**

It is affiliated to Shanghai Municipal Commission of Commerce. All signed electronic contracts will be delivered directly from our platform to the governmental institute's system which could act as an independent third party to conserve the electronic contracts. This means that all electronic contracts will be confidential and available for future contract disputes as strong electronic evidences.

### **Shanghai Electronic Certification Authority Center**

It is certified by the Shanghai Municipal Government to be the only digital certificate issuance and management authority. Moreover, the 2-dimensional bar code encryption technology used in our platform is the product owned by the CA center with copyright.

### **Information Industry Co. Ltd., Shanghai New World (Group)**

The commercial entity may act as our cooperation partner and apply the platform to their own B2B website which specializes in hardware tool transactions.

The combination of production sector, university and research sector is conducive to overall development of our platform.

## **5 Conclusion**

Contract making is an essential process of e-business activities; however, no appropriate mechanism has been developed to fill the blank of the field. Thus, our study breaks the ice and raises the innovative Network Platform for Electronic Contract Making. The study resolves the following problems through introducing the highlights, design and operation prospects of the platform:

First of all, though online shopping platforms, payment manners and post-buy services are becoming more and more sophisticated, nothing to do with contract making process. Our platform implements the whole e-business activity flow.

In the second place, the platform sets a set of mechanisms available for interactive negotiation among willing parties, breaking through the bottleneck of traditional electronic contracts, the click ones.

In the third place, the platform forms a series of electronic contract templates, summarizing almost all existing common contracts.

In the fourth place, in accordance with related electronic signature laws the innovative application of electronic signature on electronic contracts endows electronic contracts equal legal rights with paper ones.

Finally, the cooperation with Shanghai Trade Point and Shanghai Electronic Certification Authority Center guarantees the creditability and security of electronic contracts.

In conclusion, the main aim of our study is to improve the development of e-business via the pioneering application of electronic signature on electronic contracts. However, there is no garden without its weeds, including our platform. There must be some inappropriate issues in our study and we are expecting your pointing out without hesitation.

**Acknowledgments.** This paper is supported by National Natural Science Foundation of China Project (Project Number: 70973097) and China Shanghai Leading Academic Discipline Project (Project Number: S30504).

## References

1. Wang, G.-F., Jing, R.-T.: Comparison and Outlook for Researches of Negotiation Support System from Domestic & Abroad. *Value Engineering*, 9–13 (August 2004)
2. Wheeler, L.H., Wheeler, A.M.: Method & System for Using Electronic Communications for An Electronic Contract. US Patent 7, 200, 749 (2007)
3. Chieu, T.C., Nguyen, T., Maradugu, S., Kwok, T.: An Enterprise Electronic Contract Management System Based on Service-Oriented Architecture. In: *IEEE International Conference on Services Computing*, pp. 613–620 (2007)
4. Paulsen, C.A., Nguyen, B.T.: Electronic Signature Capability in a Gaming Machine. US Patent 6, 935, 951 (2005)
5. Changa, I.-C., Hwanga, H.-G., Hungb, M.-C.: Factors Affecting the Adoption of Electronic Signature: Executives' Perspective of Hospital Information Department. *Decision Support Systems* 44(1), 350–359 (2007)
6. Messing, J.H.: Electronic Signature Method. US Patent 7, 039, 805 (2006)
7. Kwok, Thao Nguyen, T., Lam, L.: A Software as a Service with Multi-tenancy Support for an Electronic Contract Management Application. In: *IEEE International Conference on Services Computing*, pp. 179–186 (July 2008)
8. Yu, X., Chai, Y.-T., Yi, L.: A Secure Model for Electronic Contract Enactment, Monitoring and Management. In: *Second International Symposium on Electronic Commerce and Security*, pp. 296–300 (January 2009)
9. Solaiman, E., Molina-Jiménez, C., Shrivastav, S.: Model Checking Correctness Properties of Electronic Contracts. In: *Orlowska, M.E., Weerawarana, S., Papazoglou, M.P., Yang, J. (eds.) ICSSOC 2003. LNCS, vol. 2910, pp. 303–318. Springer, Heidelberg (2003)*

# Implementation of the Building Decoration Design Demonstration System Based on Virtual Reality

Wu Feng, Wang Wei, and Liu Xiuluo

Beijing Special Engineering Design and Research Institute, P.R. China

**Abstract.** This paper established a demonstration system for building decoration design based on virtual reality, which uses the 3D modeling technology to construct the building scene, and the scene roaming technology to realize human-computer interaction. The system can dynamically follow the design progress and comprehensively display design efforts, so as to help designers discovering design problems in advance and avoiding the pitfalls.

**Keywords:** Virtual reality, human-computer interaction, roaming, collision detection, GPU, texture mapping, 3D modeling.

## 1 Introduction

Virtual reality[1] is a new study in the information field. It can take advantage of the computer to accurately describe the real world, and the human-computer interaction to simulatively reform the real world, with good immersion, interaction and timeliness. It is now widely used in the space military, construction, manufacturing industry, urban planning and other areas. For example, designer can use the virtual reality technology to incrementally modify the design schemes with low time cost, so that more time and energy paid on the designing. However, it must take a lot of time to render a new effect picture[2] for any modification, thus greatly reducing the design time and eventually impacting on the progress and quality. For demonstrating, designer can use the roaming technology to vary the angle of view as user wanted, to help roundly examining designs, can set the automatic roaming path to lead user feeling design achievements, and can combine 2D CAD blueprint with 3D decoration models, aiding by distance measure tool, to scientifically judge designs.

This paper established a building decoration design demonstration system based on virtual reality, aiming at truly, flexibly and roundly reflecting design achievements. Comparing with other similar systems, we make several contributions to the technique application in quest3d platform. Firstly, simplify the 3D max model's exporting flow by comprehending whose information storing structure. Secondly, find the method to revise stretching distortion in viewport transform by researching the camera space transformation technique. Lastly, realize a lightweight collision detection algorithm based on the extracted translation and rotation information from the transformation matrix.



## 2 System Designing and Implementing

### 2.1 System Framework

The building decoration design demonstration system is composed of two major modules: 3D scene and models constructing module and roaming driving functional module, showing in Fig.1. The former using 3ds max to construct the scene and models, has many sub-modules as follows: 3ds max modeling, material setting, texture mapping, light setting and shadow baking. The latter using the quest3d platform to roam and drive the whole system, has many sub-modules as follows: scene roaming, interactive controlling and information displaying.

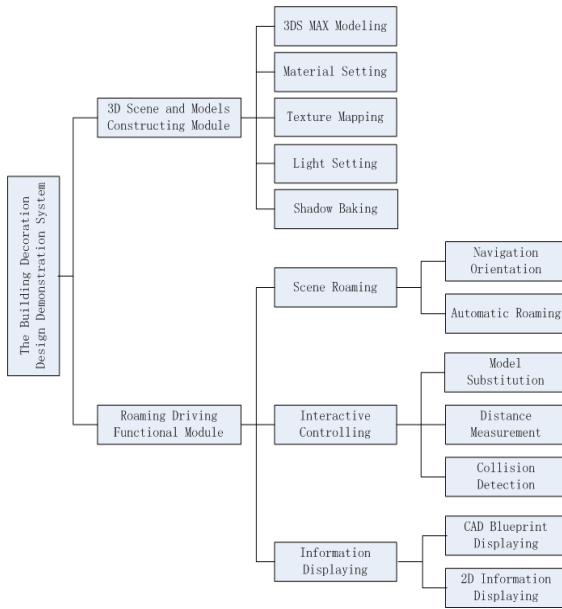


Fig. 1. The system framework

### 2.2 3D Scene and Models Constructing Module

To ensure the reality of the system, we construct the 3D models of the appearance, structure and function zoning partition of the building contrasting to CAD blueprint, in appropriate proportion in 3ds max. To enhance the immersion, use texture mapping technique[3,4], such as alpha mapping, light mapping, gloss mapping, and environment mapping, to add color, texture and shadow on models' surface. Considering the operating efficiency, reduce the amount of face in models as much as possible, and confine the size, type and quantity of texture strictly. Furthermore, only load the requisite scene or models currently to CPU. Considering the design schemes' incremental modification, split models according to the change frequency.

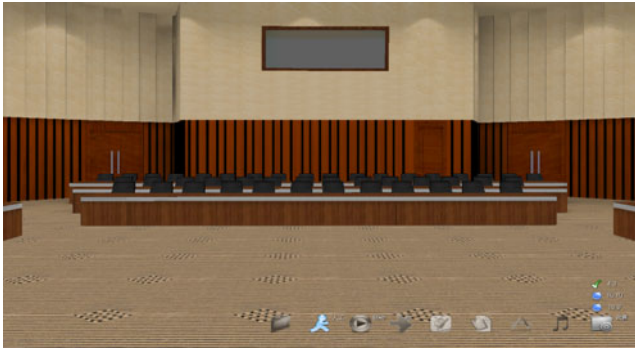
## 2.3 Roaming Driving Functional Module

### 2.3.1 Scene Roaming

It is allowed that user can 360 degrees observe the 3D scene by controlling the camera lens translating and rotating, can adjust the lens' focal length by setting the zoom value of camera, and can prevent viewport transform, which is the projection from quadrate projection window to rectangular viewport, from stretching distorting by setting the size value of camera according to aspect ratio[4], showing in Fig.2.



(A) System normally displaying when aspect ratio equals 16:10



(B) System normally displaying when aspect ratio equals 4:3

**Fig. 2.** System normally displaying under various aspect ratio

The roaming modes are classified into two categories: navigation orientation and automatic roaming. The former using navigation maps guides user entering any floor and room directly, showing in Fig.3. The latter planning a fixed roaming path and setting the wandering speed and angle velocity, allows looking through all design schemes continuously in the first person perspective. In order to achieve optimal demonstration effect, need to exhibit the grand style of decoration at the initial phase of roaming, and turn over camera lens' speed, focal length and movement in the process adaptively.

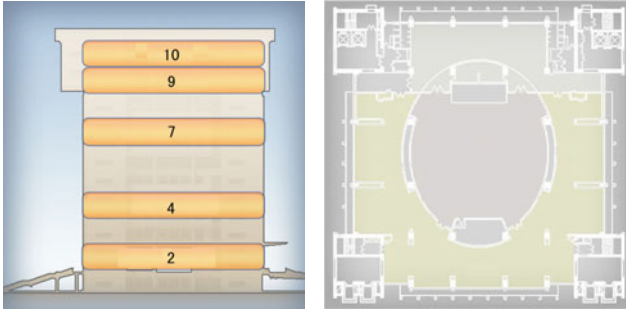


Fig. 3. The navigation orientation interface

### 2.3.2 Interactive Controlling

Interaction control module consists of model substitution, distance measurement and collision detection. Model substitution provides the switch function among all spare design schemes for users to compare and select, showing in Fig.4.



(A) The design scheme 1 for room A



(B) The design scheme 2 for room A

Fig. 4. Two spare Schemes for room A

Distance measurement provides the function at any time convenient for users to measure models, showing in Fig.5. Collision detection[3] prevents penetrating walls (or objects) or sinking into underground while roaming, is the basis technology for walkthrough camera and distance measurement, and also, it can be applied to detect the intervention between furnishings and the water, air-condition, electricity, communication equipments in the building or rooms.

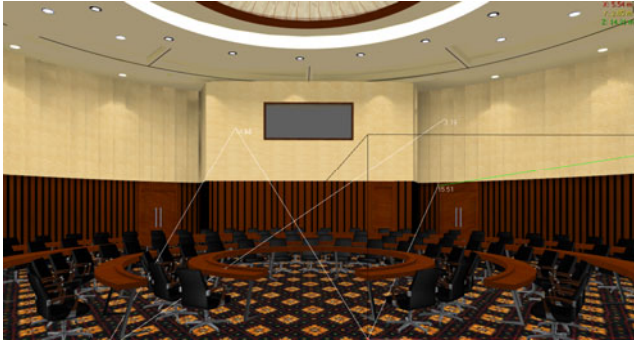


Fig. 5. The distance measure

### 2.3.3 Information Displaying

Information display consists of 2D information display and CAD blueprint display, the basic means to display the decoration design schemes, showing in Fig.6.

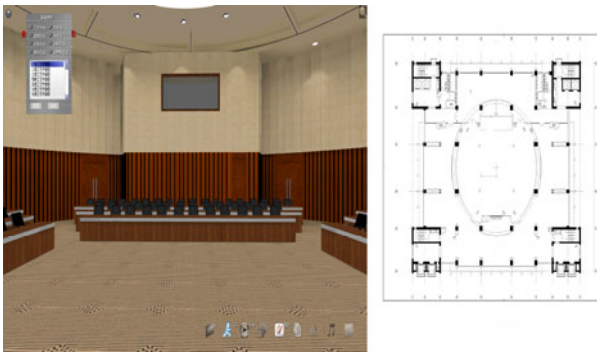


Fig. 6. The CAD blueprint displaying

## 3 Key Technologies Researching

The building decoration design demonstration system involves the following key technologies: the 3D model's information storing structure, camera space transformation and collision detection.

### 3.1 The 3D Model's Information Storing Structure

In order to effectively drive 3ds max model in quest3d platform, such as locating and rotating, we must put the translation and rotation center on the model itself. Generally, set the model's coordinate value zero to put the local origin on itself, before exporting as the first exported type. It involves many difficult, time-consuming and fallible manipulations in 3d max, so needs to find an instead method to simplify the exporting flow.

Discuss with two types of the exported 3ds max model: 3D Object and 3D ImportObject. They both have the following information storing structure[5]: Motion, containing the position, rotation and model size information, offset motion, used to adjust the local coordinate system, surface, containing the vertex data, material and texture information. The Structure is showing Fig.7[5]:

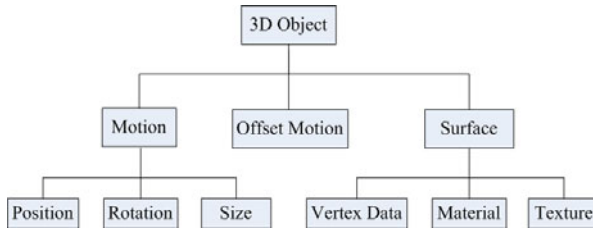


Fig. 7. The 3D model's structure in quest3d

Although have the same structure, it is quite different between them. 3D Object puts the origin of the local coordinate system on the origin of the world coordinate system, however, 3D ImportObject puts the origin of the local coordinate system on the model itself. Therefore, the former's position and rotation are both (0,0,0) generally, and it saves each vertex's offset and rotation amount to the world origin in Vertex Data. The later's position and rotation contains the local origin's offset and rotation amount to the world origin, and it saves each vertex's offset and rotation amount to the local origin in Vertex Data.

According to the analysis above, as the instead exporting method, which avoids manipulating the model's local origin in 3d max, we can either set the offset motion in quest3d to put model's local origin on itself, after directly exported as the first type, or export as the second type directly. The typical application for the 3D ImportObject is the human driving[5].

### 3.2 Camera Space Transformation

In the building decoration design demonstration system, it needs to do the camera space transformation, realized by the transformation matrix  $M$  in the transformation pipeline, to provide the scene roaming function. The matrix  $M$  shows as following[6,7]:

$$M = \begin{bmatrix} M_1 & t \\ 0 & 1 \end{bmatrix} \tag{1}$$

where  $M_1$  is a  $3 \times 3$  matrix,  $t$  is a  $3 \times 1$  offset matrix, and the 4th row vector is defined as  $(0, \dots, 0, 1)$ . To facilitate the implementation of matrix multiplication, adopt the

homogeneous coordinate to represent vertex and set the value  $w$  to 1. Generally, the transformation matrix  $M$  can be written in the form of the three matrices' product- $T$  (translation),  $R$  (rotation) and  $S$  (scale) [6,7]:

$$M = TRS = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_x & r_y & r_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where  $t = (t_x, t_y, t_z)^T$  denotes the translation part,  $r_x, r_y, r_z$  are column vectors forming the regularization axis, and  $s_x, s_y, s_z$  are scale factors related to revising the stretching distortion in viewport transform. In quest3d, the default value of  $s_x, s_y, s_z$  suits to case that the aspect ratio equals 4:3, so let  $s_x = (16/9)/(4/3)$ ,  $s_y = s_z = 1$  when aspect ratio equals 16:9.

Given the scale factor, introduce how to extract the translation and rotation information from the transformation matrix  $M$ [7]. Translation vector can be get from the last column of  $M$ . The rotating vector can be get by dividing the first three columns of  $M$  by the corresponding scale value respectively. Next, introduce how to solve the inverse of  $M$ [7]:

$$M^{-1} = (TRS)^{-1} = \begin{bmatrix} r_x^T/s_x & -r_x \cdot t/s_x \\ r_y^T/s_y & -r_y \cdot t/s_y \\ r_z^T/s_z & -r_z \cdot t/s_z \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A^{-1} & -A^{-1}t \\ 0 & 1 \end{bmatrix} \quad (3)$$

From (3), we can use the scale factor's square to calculate of the first, second and third row as following:

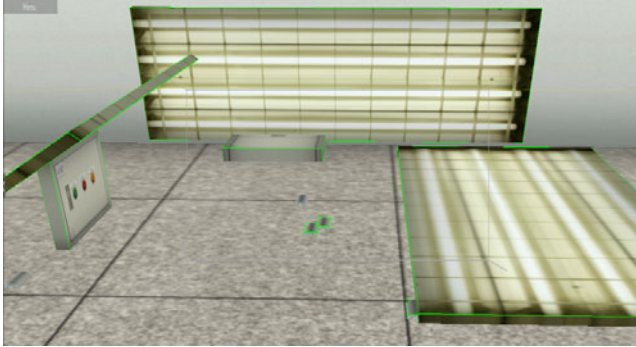
$$\left( \frac{s_x r_x}{s_x^2} \right) = \frac{r_x^T}{s_x}, \left( \frac{s_y r_y}{s_y^2} \right) = \frac{r_y^T}{s_y}, \left( \frac{s_z r_z}{s_z^2} \right) = \frac{r_z^T}{s_z} \quad (4)$$

### 3.3 Collision Detection

In the building decoration design demonstration system, the collision detection algorithm is the basis of the penetration detection between the walkthrough camera and the scene, the contact point detection in distance measure, and the interference detection in decoration disposing. The collision detection algorithms are generally divided into two phases[7]. The former finds out the candidate collision pair collection, in which any pair has high probability of collision. The latter is responsible for finding out real collision pairs in the collection by accurately intersecting test.

Next, we will introduce the application of the axially aligned bounding box *AABB* algorithm[7], for instance, to interference detection in decoration disposing. In the quest3d platform, according to 3.2 section of translation and rotation information extraction method, firstly, we can obtain the model's global translation and rotation information in the world from the motion matrix, then, compute the global size by using the global rotation to multiply the given local size, lastly, compute the global *AABB* on basis of the global translation and the global size to detect interference.

Fig.8 shows intersecting test results in the world, where green boxes represent models' local AABB.



**Fig. 8.** The AABB intersecting test in the world

## 4 Conclusion

Combined with the demand of truly, flexibly and roundly reflecting building decoration design schemes, this paper the building decoration design demonstration system, which supports precise 3D Modeling and real-time human-computer interaction, based on virtual reality. This system can dynamically follow the design progress and comprehensively display design efforts, thus do better than effect pictures in helping designers to discover design problems in advance and to avoid the pitfalls.

In future, we consider introducing cartoon's excellent demonstrating skills into the system's automatic roaming, to enhance the system's virtual and real sense.

## References

1. Donald, H., Pauline Baker, M., Wrote, Cai, S., Song, J., Cai, M.(trans.): Computer Graphics with OpenGL, 3rd edn. Publishing House of Electronics Industry (June 2005)
2. Qu, B., Zhao, Y., Zhao, Q.: Design and Implementation of a Virtual Home Ramble System Based on Virtools. Computer Engineer & Science 31(12), 130–133 (2009)
3. Akenine-Moller, T., Eric, H., Wrote, Pu, J.(trans.): Real-Time Rendering. 2nd ed. Publishing House of PKU (July 2004) (in Chinese)
4. Luna, F.D., Wrote, Fei, D.(trans.): Introduction to 3D Game Programming with DirectX 9.0. Tsinghua University Press (April 2007) (in Chinese)
5. Quest 3D Documentation
6. Fletcher, D., Ian, P., Wrote, Shi, Y., Chen, H., Wang, R.(trans.): 3D Math Primer for Graphics and Game Development. Tsinghua University Press (July 2005) (in Chinese)
7. Scott, J., Wrote, Xiang, Z., Tao, S.(trans.): Game Programming Gems 7. Posts & Telecom Press (July 2010) (in Chinese)

# Expanded Study on Criticality in Schedule Plan of Project

Li Dayun and Guo Shu-hang

School of Information, Central University of Finance and Economics Beijing, 100081, China  
lidayun@126.com, guoshuhang@hotmail.com

**Abstract.** This article mainly concerns with resource-constrained schedule risk and comes up with a method which uses Genetic Algorithm to find out the shortest time of a project and bases on Monte Carlo Simulation to estimate schedule risk. First, a new concept of is proposed, which optimizes the resource constraints and epitomizes the law of diminishing marginal utility of resource. Second, introduce how to take advantage of Genetic Algorithm and Monte Carlo Simulation in schedule-risk-management. Finally, an example manifests that it is effectual of Monte Carlo Simulation to estimate schedule risk and assist decisions in project under resource-constrained circumstances.

**Keywords:** Effective resource unit, schedule risk, Genetic Algorithm, Monte Carlo Simulation.

## 1 Introduction

In project management, all the activities inevitably possess uncertainties; therefore, there exists various risks. Schedule is one of the four objectives of Production Development management and has significant effects on the success of the whole project. Currently, in practices of project management, the network planning technology with core of classical PERT/CPM has been widely used[2][6]. But the premise of this approach is no resource constraint. If one or several resources supply is limited, it may delay the entire project. Some heuristic algorithms such as genetic algorithms can solve such problems that how to plan the project schedule and estimate schedule risk under resource constraints.

Monte Carlo method as a simulation technology has a big advantage in solving project risk issues and other complex random process. Furthermore, with the development of computer technology, a large number of samples are possible, which increase problem-solving accuracy, can good practice guidance[4][5]. This article mainly concerns with Monte Carlo-based project schedule risk in quantitative assessment methods under resource constraints. We hope to have a more rational analysis of the project duration variation, which supports project decisions.



## 2 Nomenclature

In projects, resources for program activities have effects on the duration, because the allocation of resources and whether they are used for the project play a great role in duration. To illustrate this problem, first present a few concepts of activity duration, which are activities duration, working hour and resources unit [3][5].

(1)Activities duration: the actual time of completing the task.

(2) Working hours: the working hours of mission resources.

(3)Resource units: the measurement of each resources unit’s working time within the duration, usually expressed as a percentage. In this article, the actual resource units will replace it.

Based on reality, new concept of an effective resource unit is proposed, which is the actual contribution to the output of resources, as a function of the actual resource units. It designs to balance diminishing marginal contribution of resources caused by management costs, free rider problems, when resources increase to a certain level. The function from the actual resource units to the effective resource unit is characterized by: first derivative is positive, the second derivative is negative, when the actual resource units tends to infinity, the effective resource unit has a limit, its value may be determined by the experts.

Define the function from effective resource units to the actual units as mapping function. Accordingly, the mapping function has the following characteristics: first derivative is positive, the second derivative is also positive, when the actual units approach limit, the actual resource units tends to infinity.

Relationship between working hours and effective resource unit is:

$$\text{Working Hours} = \text{Activities duration} \times \text{Effective resource unit}(1)$$

## 3 Problem Description

Resource constrained project scheduling problem is a prerequisite to estimate resource constraints schedule risk. The resource-constrained project scheduling problem can be described as below: the project contains a collection of activities  $V=\{0,1,\dots,V,V+1\}$ , in which  $V_0$  and  $V_{V+1}$  are the virtual activity, representing the start and finish date of the project.  $L$  kinds of renewable resources are needed to complete the project. Resource  $l \in L$  needs  $R_w(w=1,2,\dots,W)$  resources during the execution of the project. For the task  $v \in V$ , duration is  $D_v$  and  $d_0=d_{V+1}=0$ . To complete the task  $v$ , requirement of resource  $w$  is  $r_{v,w}(w=1,2,\dots,W)$  and  $r_{0,w}=r_{V+1,w}=0$ .  $d_v$ ,  $r_{v,w}$  and  $R_w$  are positive integers.

The project process must satisfy two kinds of constraints:

(1) Topology constraints.

Assume  $P_v$  is the predecessor activity set of  $v$ . For any activity  $u \in P_v$ , if  $u$  has not finished, the activity  $v$  should not be started.

(2) Resource constraints.

During the implementation of the project at any time, each resource of the activity being performed cannot exceed the total amount of that kind of resource.

The goal is to determine the start and finish date of each activity and to minimize project duration within time constraints and resource constraints.

The mathematical model can be described as below:

$$\text{Min } F_{V+1} \tag{2}$$

$$F_v \leq F_v - d_v \quad v = 1, 2, \dots, V, V + 1; p \in P_v \tag{3}$$

$$\sum_{v \in A(t)} r_{v,w} \leq R_w \quad w \in W; t \geq 0 \tag{4}$$

$$F_v \geq 0 \quad v = 1, 2, \dots, V + 1 \tag{5}$$

In the formulae above,  $F_v$  stands for completion time of activity  $v$ .  $A(t) = \{v \in V | F_v - d_v \leq t < F_v\}$  is a collection of activities being performed at the time of  $t$ . The completion time of schedule  $S$  is  $(F_1, F_2, \dots, F_V)$ . Function (2) is the objective function which minimizes total project duration. Formula (3) and formula (4) presents topology constraints and resource constraints respectively. In the model, equation (4) cannot be determined before achieving optimal value. So the model is only a conceptual model which could not be directly used to solve.

In the following parts, the article uses Genetic Algorithm to solve resource-constrained activities arrangement problem. That is to say, a single project scheduling under certain topology constraints and resource constraints is resolved. After all, resource-constrained schedule risk is a probabilistic problem. Put another way, how possible the project can be completed within the stipulated time or how possible cannot be completed within the stipulated time. That is the schedule risk which requires statistics based on many times of simulations. This article uses Monte Carlo Simulation approach to estimate schedule risk.

## 4 Genetic Algorithm

The genetic algorithm optimization features can easily achieve a resource-constrained project scheduling optimization. The risk of the project could be measured by shortest duration under resource constraints.

### 4.1 Design of Encoding and Decoding

A priority-based sequence of activities encoding is used in the coding[1], as described in Figure 1. The project activities not only should follow resource constraints, but also follow topology constraints. And the priority-based sequence of activities encoding not only reflects the random execution of activities, but also follows the topology constraints. That is, when there are multiple activities whose predecessor activities have finished and resource constraints are fulfilled, activity with maximum priority will be implemented firstly. Thus arrangement of activities is confirmed.

Chromosome structure is formed by the task priority. The value  $Pr_i$  is used to indicate the priority of activity  $i$ . It is a real number between  $[0, 1]$ . The greater the value is, the higher the priority is.

1	2	3	.....	n	Activity Num
$Pr_0$	$Pr_1$	$Pr_2$	.....	$Pr_j$	Priority

Fig. 1. Chromosome encoding

Divide decoding chromosome into two steps:

(1)Arrangement of activities is generated through arranging the best chromosome. Define unfinished activities whose precursor activities have been completed as eligible activities. Therefore, each activity must be one of mutually exclusive three states: completed activities, eligible activities, and other activities.

(2) The ultimate shortest duration is calculated according to the arrangement of activities and resource constraints.

Decoding process is described as Figure 2:

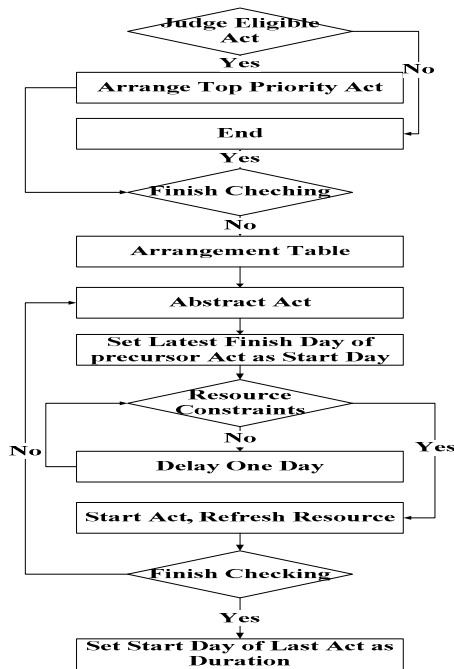


Fig. 2. Decoding flow chart (Act stands for Activity)

### 4.2 The Fitness Function

Calculating the shortest duration of activity is a minimum problem. So we should divert schedule into fitness value to ensure selected individuals has the large fitness value. S[16]et  $I$  to be the current population of the  $i$ -th chromosome. Then the fitness function [9] is:

$$fitness(I) = D_{max} - D_i + \gamma \tag{6}$$

In the function (6),  $D_{max}$  is the maximum duration of the current population.  $D_i$  is the schedule of current individual.  $\gamma$  is a positive real number of the open interval (0,1). Using  $\gamma$  could transform selection behavior from fitness value proportional selection to pure random selection.

### 4.3 Design of Genetic Operators

(1) Design of selection operator

According to the fitness value calculated in the above steps, use roulette selection operation.

(2) Design of crossover operator

Use single-point crossover operator to complete the operation.

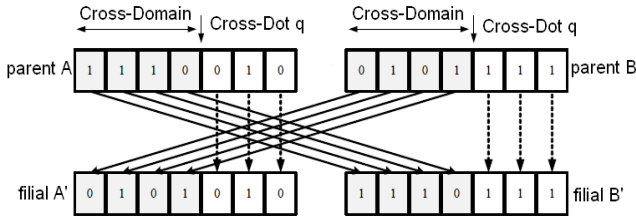


Fig. 3. Single-point crossover operation

(3) Design of mutation operator

Use single uniform mutation operator to complete the operation.

(4) Design of genetic parameters

The results showed that genetic algorithm can obtain better performance when the crossover probability is 0.4 to 0.6 and mutation probability of 0.03 to 0.05 [1].

## 5 Monte Carlo Simulation

Monte Carlo method, known as a random sampling technique or a statistical testing method, is a method to achieve approximate solutions of mathematical, physical, engineering and technical problems through statistical analysis of random variables and stochastic simulation. It is based on statistical probability theory[2][4]. Monte Carlo method is to simulate specific problems by using sampling sequence of

different distributions of random variables and find estimated value of the solutions to the problem.

### 5.1 Sampling

The effective resource unit obeyed triangular distribution according to experience. Experts can give three kinds of estimation: the minimum  $a$ , the most likely value  $m$  and the maximum  $b$ . Sampling process can be described as below: generate a stochastic real number between (0,1).

- (1) Determine whether  $R$  is greater than  $\frac{m-a}{b-a}$ ;
- (2) if  $R \leq \frac{m-a}{b-a}$ ,  $x = a + \sqrt{(m-a)(b-a)R}$ ;
- (3) if  $R > \frac{m-a}{b-a}$ ,  $x = b - \sqrt{(b-m)(b-a)(1-R)}$ ;

According to above algorithm, a random number of triangular distribution is obtained.

In practice, working hours are often fixed in a specific activity. For each activity, sample effective resource unit, and according to the formula (1), calculate the duration of each activity:

$$\text{Working Hours} = \text{Activities duration} \times \text{Effective resource unit}(1)$$

### 5.2 The Number of Simulation

Steps of determining the number of stimulation can be described as below[8]:

- (1) Determine the degree of confidence  $\beta$ , and then calculate  $\Phi(U) = \beta + 1/2$ . Check  $U$ -value from the look-up table;
- (2) Determine the error  $\varepsilon$  that can be accepted;
- (3) Run simulation  $n_0$  ( $n_0 \geq 30$ ) times to calculate the sample variance  $S^2$ ;
- (4) Take  $\varepsilon, U, S^2$  into  $N = \frac{U^2}{\varepsilon^2} S^2$ , the number of stimulation  $N$  is achieved.

### 5.3 Mapping Function

Mapping function could be determined either by the actual situation, or by the following simple methods.

Assume  $y$  is the actual resource unit,  $x$  is effective resource unit,  $c$  is the effective limit resource unit. And  $c$  is greater than maximum  $b$  mentioned in the Sampling section. Here offers two ways for reference:

$$1 \ y = \frac{1}{1 - \frac{x}{c}} + (1 - \frac{x}{c}) \quad (c \geq x \geq 0) \tag{7}$$

$$2 \ y = x^n \quad (n > 1, c \geq x \geq 0) \tag{8}$$

Equation (7) is a stimulation of the effective resource unit from 0 to its limit, while equation (8) is a stimulation of the effective resource unit within sampling value range.

In practice, the mapping function of each activity can be the same or different. So does the mapping function of different resource. Manager could decide according to the specific goal of the project.

### 5.4 Monte Carlo Simulation Process

Resource-constrained Monte Carlo simulation can be described as below, shown in Figure 4:

- (1) Take effective resource unit for stochastic parameter;
- (2) Determine the number of simulation N, in order to meet accuracy requirements;
- (3) Sample effective resource unit;
- (4) Calculate the actual resource unit according to the mapping function;
- (5) Calculate the duration of each activity according to working hour and equation (1);
- (6) Record duration and the key event information;
- (7) After N consecutive samples, statistically analyze the data and draw statistical feature figures.

After the above steps, we can get a lot of stimulation data, whereby it helps estimate project scheduling risk and assist decision-making.

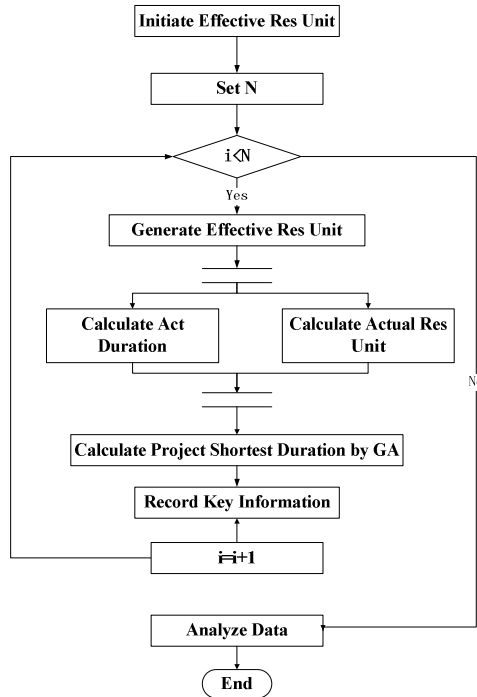


Fig. 4. Monte Carlo simulation process (Act stands for Activity, and Res stands for Resource)

## 6 Case Study

### Research Purpose

How effective project risk assessment to support decision-making has been a hot issue in academia. In order to prove that the Monte Carlo is an effective method for project risk, especially the risk assessment and decision-making in resource-constrained project. Data are abstracted from the relevant reference to conduct Monte Carlo simulation.

## 7 Data Preparation

Knitting machine is the major equipment of Sweater production. A company has developed a newly multi-functional computerized -knitting machine which already has the basic functions of automatic one [2]. The design of such knitting machine implementation of several key stages is described in Table 1.

**Table 1.** Knitting machine design phase of the implementation (a, b, m refer to minimum, maximum and most likely resource required)

Num	Precursor Activity	Working Hours	a	b	m
1	-	0	0	0	0
2	1	15	4	6	5
3	1	20	4	6	5
4	1	9	2	5	3
5	2	10	1	3	2
6	3	18	2	5	3
7	2,4	16	3	6	4
8	3	15	3	7	5
9	6	32	3	5	4
10	5	12	2	4	3
11	5,7	6	2	5	3
12	6, 8	6	2	4	3
13	9	2	1	2	2
14	9	15	3	6	5
15	10,11	3	1	2	2
16	12	4	1	3	2
17	14	6	1	3	2
18	16	25	4	6	5
19	14,15	12	3	6	4
20	13	6	2	5	3
21	18,20	4	1	4	2
22	17	3	1	3	2
23	17,19,20	2	1	3	2
24	18	15	2	5	3
25	21	2	1	3	2
26	22	6	1	3	2
27	22,23,24,25	3	0	2	1
28	26,27	0	0	0	0

Under resource constraints of 10, the company hopes to finish within 40 days and want to know that how many days will it take with completing probability of 90%. To evaluate the expected duration, reduce the schedule risk, and provide a scientific basis for scheduling, this case gives working hours and required resource of each activity according to expert opinion. Use Monte Carlo method to stimulate the company's actual progress of the implementation[7].

The project activity network diagram is displayed as Figure 5:

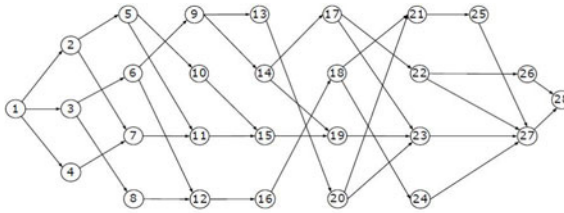


Fig. 5. Knitting machine design activity network diagram

### 8 Experiment Process

The entire process of resource-constrained Monte Carlo simulation is summarized as below, take Matlab for example:

- (1) Create a matrix, sample effective resource stochastic variables of triangular distribution;
- (2) Set effective resource unit for  $x$ , the actual resource units for  $y$ . Assume that the mapping function is  $y = x^{3/2}$ . That is 3/2 power of effective resource unit;
- (3) Calculate the duration of each activity according to working hours;
- (4) Set the degree of confidence  $\beta=0.95$ ,  $U = 1.96$  which is checked in table, set the error  $\varepsilon=0.1$ ;
- (5) Run  $n_0$  ( $n_0=30$ ) times of stimulation to calculate the sample variance  $S^2=25.2057$ . Take  $\varepsilon$ ,  $U$ ,  $S^2$  into  $N = \frac{U^2}{\varepsilon^2} S^2 = 9683.021712$ . The number of simulation  $N$  is 10,000;
- (6) Simulate 10,000 times of duration through genetic toolbox in Matlab.

### 9 Evaluation of Results

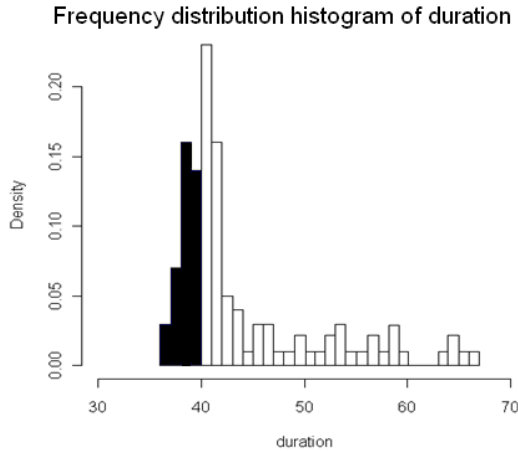
Set the number of simulation to 10,000, analyze project duration of the simulated data, the results are listed below:

Table 2. Monte Carlo simulation results

Mean	Variance	1st Qu.	3rd Qu.	Min	Max
41.75	20.43182	39	42	36	67

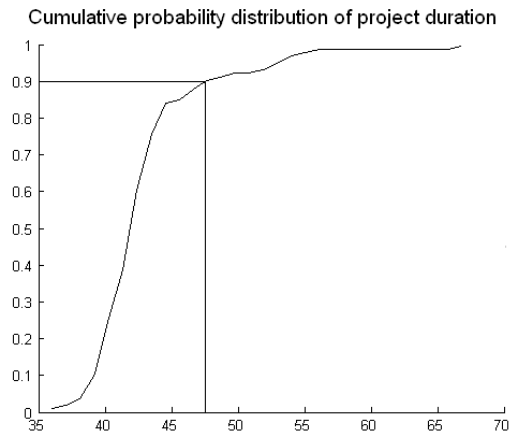


Data show that the maximum duration is 67 days, the minimum duration is 36 days, the mean duration is 41.75 days, and simulation variance is 20.43182. The frequency distribution of duration is shown in Figure 6:



**Fig. 6.** Frequency distribution histogram of duration

From Figure 7 cumulative probability distribution of project, we can conclude that the probability of the project completing within 40 days under resource constraints is 24.528%. And 45 days are needed to complete the project with 90% probability of completion, we can see while 40 days are too risky. The project manager must modify project schedule or enhance risk management to ensure smooth implementation of the project.



**Fig. 7.** Cumulative probability distribution of project duration

However, we must figure out which activities affect the duration to take appropriate measures. Therefore, further analysis is needed on how much activities influence on the duration.

First, identify the critical path, which is shown in Figure 8:

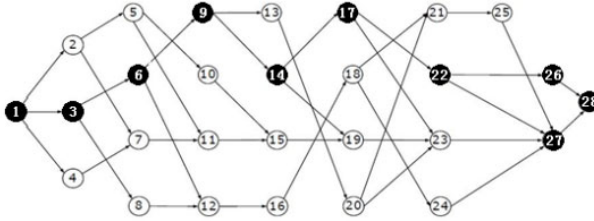


Fig. 8. Knitting machine design critical path (black activities are key activities)

To verify the critical path's impact on the total duration, multi-factor analysis of variance of 26 real activities is conducted. The results are shown in Table 3:

Table 3. Analysis of multi-factor duration variance

	Df	SumSq	MeanSq	Fvalue	Pr(>F)	
A2	1	0.73	0.73	0.4583	0.500569	
A3	1	21.7	21.7	13.5524	0.0004416	***
A4	1	7.09	7.09	4.43	0.0387553	*
A5	1	44.9	44.87	28.0248	1.21E-06	***
A6	1	26.6	26.55	16.5851	0.0001167	***
A7	1	2.08	2.08	1.3022	0.2575438	
A8	1	0.1	0.1	0.0632	0.8022725	
A9	1	40.5	40.48	25.2844	3.42E-06	***
A10	1	3.28	3.28	2.0484	0.1566321	
A11	1	6.83	6.83	4.2636	0.0424899	*
A12	1	0.51	0.51	0.3182	0.57441	
A13	1	2.21	2.21	1.3816	0.2436494	
A14	1	25.8	25.83	16.1309	0.0001419	***
A15	1	0.67	0.67	0.4214	0.5183004	
A16	1	12.8	12.82	8.0062	0.0060163	**
A17	1	4.06	4.06	2.5372	0.115513	
A18	1	3.37	3.37	2.1018	0.1514134	
A19	1	2.52	2.52	1.5751	0.2134724	
A20	1	5.52	5.52	3.445	0.0674813	.
A21	1	15.4	15.35	9.5877	0.0027786	**
A22	1	77	77.02	48.1068	1.37E-09	***
A23	1	26.3	26.29	16.4194	0.0001253	***
A24	1	40.3	40.3	25.1686	3.57E-06	***
A25	1	6.55	6.55	4.0887	0.0468372	*
A26	1	8.97	8.97	5.6	0.0206153	*
A27	1	1520	1520	949.4384	< 2.2e-16	***
Residuals		73	116.9	1.6		

Signif.codes:0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

From Table 3, activities 3,4,5,6,9,11,14,16,21,22,23,24,25,26,27 analysis of variance is more significant. That is to say, the activities have a significant effect on total duration of the project. So any delay on these activities is likely to delay the project schedule. To further prove this effect, conduct correlation test respectively. The results are shown in Table 4.

**Table 4.** The correlation test coefficient table

	<b>correlation</b>	<b>t-value</b>	<b>p-value</b>
A3	0.1035732	1.0309	0.3051
A4	0.06386812	0.6336	0.5278
A5	0.144258	1.4432	0.1522
A6	0.1108958	1.1046	0.272
A9	0.1623041	1.6283	0.1067
A11	0.09974047	0.9923	0.3235
A14	0.1221733	1.2186	0.2259
A16	0.1091952	1.0875	0.2795
A21	0.08087348	0.8032	0.4238
A22	0.1834439	1.8474	0.06771
A23	0.1059887	1.0552	0.2939
A24	0.1305957	1.304	0.1953
A25	0.05288665	0.5243	0.6013
A26	0.02580594	0.2556	0.7988
A27	0.9591201	33.5506	< 2.2e-16

Table 4 shows that activity 27 has a very high correlation coefficient. So the best way to reduce the duration of the project is to focus on deep-rooted factors that cause changes in activity 27. And increasing the associated allocation of resources, reducing resource constraints, or extending the duration of the project should also be considered. Thus the whole project could be completed within a larger probability and thereby reduce the schedule risk.

## 10 Conclusion

Project management schedule risk assessment must take full account of the impacts that various resource constraints may exert on the project, in reality. In reality, the schedule of the project is undoubtedly the core of project management. Therefore, it is critical to consider various possibilities of resources. This paper introduces the concept of effective resource unit not only to optimize the resource constraints but also to reflect diminishing marginal utility of resources. Also, effective resource units are used as random variables in the article. A large number of simulation models of duration can be obtained through Monte Carlo simulation. The distribution of these statistics can help to analyze the project schedule risk. In the actual situations, managers should try to estimate resource-constrained schedule risk as early as possible. Increasing resources or duration of the project could serve as means to predict and then avoid risk.

**Acknowledgments.** 211 project self-sponsor.

## References

1. Guang, P., Chen, Y.-W., Hao, W.: Software project risk management theory and methods of Research. *Control and Decision* (2007)
2. Zhong, D.-H., Liu, K.J., Yang, X.-G.: Construction schedule uncertainty of network simulation. *Systems Engineering Theory and Practice* (2) (2005)
3. Jiang, G., Chen, Y.-W.: Based on the progress of critical chain project risk management software *Computer Applications* (1) (2005)
4. Chen, C.-C., Qi, H., Di, P.: In Monte Carlo simulation of the risk assessment system design and implementation of the project management techniques (7) (2010)
5. Guo, Q., Xu, G.-Y., Xing, J.: Random schedule risk analysis of network modeling and simulation. *Armored Force Engineering Institute* (01) (2005)
6. Chen, W., Li, B., Shi, Y., Teng, H.: Solving RCPSp problem with the distribution of the estimated differential evolution algorithm. *Computer Engineering and Applications* (2011)
7. Yang, L.: Based on genetic algorithm-based project scheduling resource-constrained optimization problem. Shanghai Jiaotong University (2007)
8. Wang, X., Li, Y., Zhang, K.: Based on genetic algorithm multi-mode resource-constrained project scheduling. *Computer Science* (2006)
9. Tang, A.-G., Wang, R.: Based on Bayesian network model of software project risk assessment. *Computer Engineering* (22) (2008)

# WCF Applications in RFID-Based Factory Logistics Management System

Zhang Jianqi, Li Changzheng, and Bao Fumin

Xi'an Aerospace Automation Corp., Ltd., Xi'an, Shaanxi, China  
zhangjianqi1975@163.com

**Abstract.** By studying the distributed environment of RFID-based factory logistics management system and aiming at meeting the requirements of its security and reliability, this paper proposes a specific WCF Service Application solution. With the application of the WCF service model to the RFID-based factory logistics management system, the interaction problems among the device management, transaction processing and data services are solved effectively and one well application effect is obtained. Besides, the usage of MVC architecture separates the implementation and presentation, provides data protection at the same time and greatly enhances the security and reliability of the system.

**Keywords:** WCF, RFID, Logistics management system, Distributed, MVC.

## 1 Introduction

WCF (Windows Communication Foundation) is an application programming interface of data communication which is developed by Microsoft [1]. As a part of .NET framework, it is one of the four main system frameworks and the rest three are Windows Presentation Foundation, Windows Workflow Foundation and Windows Card Space. Radio Frequency Identification (RFID) is a non-contact automatic identification technology which can be widely applied to various fields such as logistics management, manufacturing, aviation and military and identification fields, etc. By combining high reliability and security of WCF technology with natural advantages of RFID technology, this paper conducts deep research on logistics management system of the factory.

## 2 Introduction of WCF Technology

### 2.1 New Features of WCF

As a distributed programming platform based on SOAP message, WCF offers excellent security and reliability assurance to us. The application of these characteristics of WCF can make sure the reliable operation of RFID factory logistics management system.

### Security

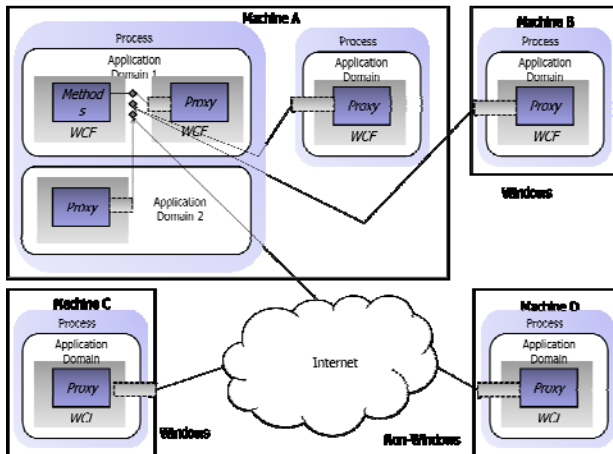
Based on the present security infrastructure and the verified security standard of SOAP message, WCF provides common platform which is interoperable and for safe information exchange. The security mechanism of WCF lies in two aspects: WCF Authentication Mechanism and WCF Transport Security Mode.

### Reliability

WCF's main differences from other service-oriented technologies are its Transport Reliability and Message Reliability.

## 2.2 System Architecture of WCF Technology

As a standard platform of Microsoft's next generation of distributed application and Service-Oriented Architecture (SOA), WCF integrated the present distributed system technologies and formed a unified and efficient service development platform [2]. Even invoking the service of local machine's internal storage, WCF does not allow the direct service interaction between clients. On the contrary, the client always uses proxy to transmit invocation to service [3]. The public operation and service of proxy are identical and at the same time, some methods of proxy management are added. WCF allows the client to cross executive boundary to correspond with service [4]. In the same machine, the client can invoke the service of the same application domain; also can invoke that of the different application domains in the same process. Even, it can invoke cross the process, cross the machine border as well as cross platforms (as is shown in Fig.1)



**Fig. 1.** WCF applications call across domains, cross-process, cross-border machine, cross-platform

The security and reliability offered by WCF and its characteristics of across domains, cross-process, cross-border machine, and cross-platform can fully satisfy the design requirements of RFID-based factory logistics management system and can insure the technique data all round.

### 3 Introduction of RFID Technology

RFID is a technology which delivers messages through RF signal space coupling (alternating magnetic field or electromagnetic field) so as to automatically identify the target objects and obtain related data at the same. As for this technology, its identification work can be conducted without human intervention and it can be applied to any severe environment. With fast and convenient operation, RFID technology can identify high-speed moving objects as well as many labels in the meantime [5].

The complete RFID system consists of three parts, i.e. Reader (or Interrogator), Transponder (or Electronic Tag) and application software system. Its operating principle is as follows: Reader sends a specific frequency of radio waves; Transponder receives RF signal sent by Reader after it enters into magnetic field, then relying on the obtained energy of induction current, it sends the message stored in chip; at this moment, Reader receives and decodes data in order and then sends them to application to receive corresponding treatment. System hardware structure is shown in Fig. 2 as follows:

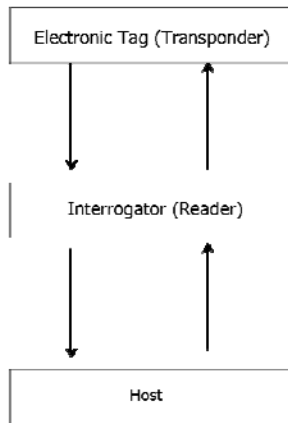


Fig. 2. RFID system structure

The working frequency of RFID ranges from low frequency, high frequency to ultrahigh frequency. Thus, it can work in the typical radio bands such as 123KHz, 433MHz, 800/900MHz and 2.45GHz [6]. Because the working frequency is different, the required technique data are in diversity too. In RFID factory logistics management system, ultrahigh frequency is used: the frequency ranges from 902 to 928 MHz; and it is adjustable. What's more, in this system, read-write distance is longer than 3 meters; it can identify 50 cards at least simultaneously and read access time is shorter than 3 seconds. Such design requirements can be fully satisfied by RFID.

At present, the application of RFID technology is basically mature. Its typical application fields includes: automation control of products process which is mainly applied to automation production line of large factory; logistics and storage automation management which are mainly used by large logistics and storage enterprise.

## 4 The Application of WFC to Factory Logistics Management System

The main research directions of RFID-based factory logistics management system includes: RFID-based logistics perception, RFID-based production logistics data collection and RFID-based inventory management [7].

### System Architecture of RFID-Based Factory Logistics Management System

Readers chosen in this system all support Ethernet Communication and the radio frequency between Reader and Electronic Tag is adjustable. Thus, it can avoid frequency overlap with outside ones. During the implementation, for those hand-held Readers, the wiring is inconvenient, but the Wi-Fi base station can be set up. The system architecture is shown in Fig. 3 as follows:

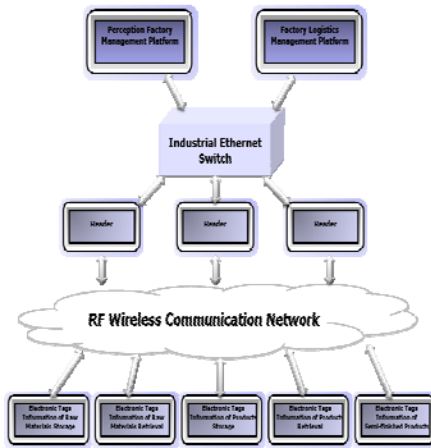


Fig. 3. RFID-based logistics management system architecture

According to practical application requirement, the RFID-based factory logistics management system can be divided into three sections among which each section exists as a subsystem, they are: logistics perception management subsystem of factory raw material, RFID-based production logistics data collection subsystem and RFID-based inventory management subsystem. Each subsystem not only individually finishes established tasks and functions, but also interacts with each other thus together realizes the practical application in the main system framework and makes sure the safe and reliable operation of the RFID-based factory logistics management system.

### Logistics Perception Management Subsystem of Factory Raw Material

The logistics perception management subsystem of factory raw material is mainly responsible for raw material identification and products identification.



The former includes: for those identified by its suppliers, this system only provides recognition; for those unidentified ones, they must first be identified and then recognized, then, combining with detailed and timely requirements provided in the production process, this subsystem provides the feedback of relevant information of each property. In the manufacturing link, the application of Sensor Networks technology can accomplish the operation of automation production line, realize the recognition and tracks of raw materials, parts, semi-finished products and products on the entire production line, reduce manual identification cost and error rate and finally promote efficiency and benefits. Adopting the IOT technology, it can quickly and exactly find the needed raw materials and parts out of various inventories through identifying the electronic tags, help managers send timely replenishment information according to production schedule to realize the balanced and stable production and in the meantime enhance the control and tracks of products' qualities.

Factory products identification is realized through electronic tags which make products as a point in the IOT and realizes the informationization of logistics and sales links. Through perception factory information processing system, it learns the products' sales conditions after leaving the factory, studies whether products' properties can satisfy customers or not and finally guides the renovation of products and improvement of production technology. Meanwhile, demanded quantity of products is predicted and reasonable production plans are formulated.

### **RFID-Based Production Logistics Data Collection Subsystem**

The RFID-based production logistics data collection subsystem is mainly responsible for the work sites data collection quality test and production inspection work.

Work sites data collection mainly refers to the collection of some quality data. The effects of manufacturing process on products quality are the greatest one and the data need to be collected are in a large scale. Whether the work sites data can be collected efficiently and accurately exerts decisive effects on enterprise's establishment of sound production records. At this stage, the data to be collected can be divided into 5 parts: person, machine, object, method and link. That is to say when, by whom, what method adopted, which machine used, in which position to put what parts on what machines. Generally speaking, the quantity of parts and products is large. So, the adoption of RFID technology not only needs the accurate and fast collected data, but also makes sure there is no increase of labor capacity. Otherwise, production efficiency is reduced and even workers have defensiveness during the process of using RFID technology. Such results backfire. To solve this problem, it should use different RFID equipment according to different manufacturing environment and working process and try its best to optimize working process.

Quality test plays an important part in the maintenance and promotion of enterprise's products quality. The quality test information collected by system mainly include: whom, when, where, what method, what products and test results.

Production inspection means that at set intervals, some factors such as production status, equipment operation status and operator manipulating status, etc. must be checked to make sure the normal operation of production. The traditional form-filling method of production inspection data collection is inefficient, with high error rate, has the problems of omitting and fabricating the data. In this circumstance, the usage of RFID Mobile Data Terminal and hand-held reader can solve this issue.

### **RFID-Based Inventory Management Subsystem**

The RFID-based inventory management subsystem is mainly responsible for storage, retrieval, inventory checkup, tracks and management of goods in progress and real time monitoring of warehouse equipment, etc.

Works of storage and retrieval includes: when the goods with electronic tags arrives at the warehouse, the reader at the entrance will automatically recognize the tags, no manual scanning is needed. In line with the achieved information, the management system will automatically update inventory lists. This procedure greatly simplifies the checkup storage program, omitting many jobs that need manpower such as trivial checking, recording and counting, etc. Retrieval is similar with this one which can realize automatic operation.

There is no need to use manual checking or bar code scanning when conducting the inventory checking. Thus, its working capacity and unnecessary loss are greatly reduced, providing larger degree of freedom for manager in the aspect of inventory control.

As for the tracks and management of goods in progress, it refers to the planning, coordinating and control of goods in progress. The usage of RFID technology can automatically collect the large amount of real-time data during production process and realize data's real-time treatment with the further help of networks. What's more, it makes the enterprise easily manage production data, promotes the products quality, reduces and avoids overstock of semi-finished products and thus improve efficiency of enterprise.

Real-time monitoring of warehouse equipment refers to confirm its position and present status through RFID reader. It is convenient for the selection of goods allocation and lines of handling tools after the goods entered the warehouse. At the same time, it can improve storage efficiency and reduce equipment's operation costs.

These three subsystems work together to accomplish the function of RFID-based factory logistics management system.

### **The Realization of Functions of WCF Service Software**

The software of RFID-based factory logistics management system adopts distributed framework. According to the different functions and levels, it is composed of three sections, i.e. server software (database software included), client software and driver software. To separate the brands of server function and that of hardware equipment, it adopts modularity pattern combination to be convenient for brand exchange of subsystems: seamless exchange realized only by the alteration of driver allocation while not depending on equipment hardware itself. Hardware exchange can not change the operation mode program code of platform. It just needs the exchange of the corresponding driver [8]. For distributed system, WCF has nature advantage. WCF itself is beyond the distributed environment base of framework [9]. No matter which program, local application program, remote application program or the application program in the Non-Windows platform, ask WCF to offer service, it treats in the same way: to provide services through the interaction of proxy and server terminal. Fig.1 fully displays the advantage of WCF in the distributed platform.

## The Adoption of MVC Three-Layer Framework

As a design mode, MVC compulsorily separate the input, treatment and output of the application program. By using MVC, such application program is divided into three core components: model, view and controller which deal with their own tasks [10], seeing Fig.4.

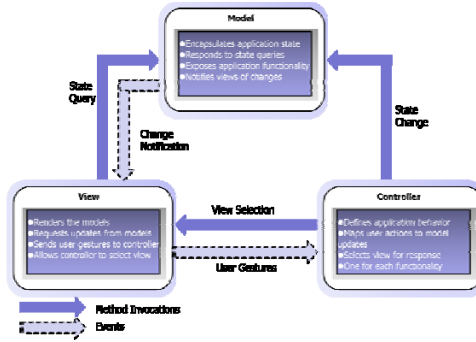


Fig. 4. MVC design pattern framework

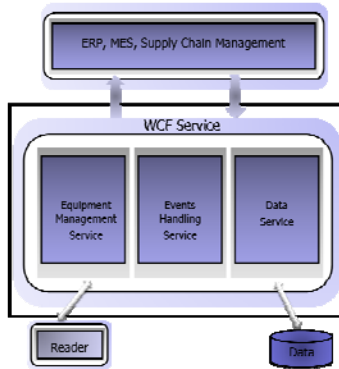
As for C/S mode, the view part is mainly realized in the client. It accomplishes the display in different environments and provides beautiful interface, fluent operation and favorable experience for the clients.

The model and controller are realized in the server terminal and are the independent one. As the core of the whole platform, server software is placed in the core server. It is responsible for information storage, treatment and screening of factory logistics system. Developed by WCF, server software uses TCP protocol to realize the reliable link on the internet and make sure the correctness of data transmission. Besides, the usage of two-way TCP blinding makes the data transmission use asynchronous mode to save network bandwidth to the largest extent and realize two-way data transmission.

Using standard MVC three-layer framework to develop application, the data base is not open to the outside which adds security of system. The adoption of unified account reduces the costs of data base software and satisfies the requirements of customers as well.

## WCF Service Module Implementation

The implementation of WCF service module on the module and controller in MVC framework operates on the server terminal and is a core part of server software. What's more, it integrates the functions of business logic layer and data access layer. WCF service module provides three services, i.e. equipment management service, event handling service and data service. Just as is shown in Fig.5:



**Fig. 5.** WCF service module implementation

Equipment management service module offers service supports for RF reader. It allows the register of reader, sending and receiving of orders and transmission of data.

As a core part of WCF service module, the event handling service module provides service interface for logistics perception management subsystem, RFID-based production logistics data collection subsystem and RFID-based inventory management subsystem of factory raw material. Besides, it offers service interface for other upper layer application software. Meanwhile, it cooperates with equipment management service module and data service module, finishing operation request task and providing services.

Data service module exists as an independent module. It dedicates to the database operation and offers reliable data service for event handling service module.

## 5 Conclusion

Combining RFID transmission characteristic with WCF's high security and reliability and using its advantages, this paper applies WCF service to the RFID-based factory logistics management system. The excellent distributed characteristics of WCF greatly promote the property of factory logistics management system. At the meantime, WCF service provides favorable guarantee for system's security and reliability. The accomplished WCF service module not only offers interface for RFID reader, but also provides a unified interface for upper layer service module. In a word, it obtains excellent application effects in the RFID-based factory management system.

## References

1. Lowy, J.: Programming WCF Services. O'Reilly Media. Inc. (2007)
2. Yan, B.: WCF-Based Distributed Application Development. *Computer Knowledge and Technology* (3), 68–70 (2008)
3. Hu, X., Li, L.: The Research and Design of WCF-Based Information System Structural Model and System Architecture. *Computer Knowledge and Technology* (2009)

4. Han, X.: WCF Service Host Remove Project Design. *Computer and Information Technology*, 12–14 (2010)
5. Jiang, H., Zhang, C., Lin, J.: *Wireless RFID Technology and Its Application and Development Tendency*. Electronic Technique Application (2005)
6. Zhang, Y., et al.: The Research on WCF-Based RFID Logistic Middleware in Modern Logistics. *Logistics Technology* (October 2010)
7. Ning, H., et al.: The Research on Chinese IOT Information Service. *Journal of Electronics* (2006)
8. Wu, X., Ma, R., Li, X.: *RFID System and Its Application in the Library*. *The Libraries* (1) (2005)
9. Zha, X., Yan, Y.: *Logistics Information System*. University of Finance and Economics Press, Dalian (2005)
10. Zhao, L.: *The Design and Realization of WCF-Based Three-layer Logistics Solution*. East China Normal University (2008)

# Research on Construction Method of Multimedia Semantic Model for Intangible Cultural Heritage

Guoxin Tan<sup>1</sup>, Tinglei Hao<sup>1</sup>, Shaoquan Liang<sup>2</sup>, and Zezhao Lin<sup>2</sup>

<sup>1</sup> National Engineering Research Center for E-Learning,  
Huazhong Normal University, Wuhan, China

<sup>2</sup> Zhongshan Golden Dragon Amusement Equipment Co., Ltd. Zhongshan, China  
gxtan@mail.ccnu.edu.cn, sht1999@163.com,  
lsq@zsdragonworld.com, linzez@126.com

**Abstract.** Considering that knowledge of intangible cultural heritage includes accumulated concealed experiences in the long term, it is difficult to completely represent the complex semantic relations among different knowledge in the form of text, graphics, images, video, 3D model, motion or other multimedia types. The digitized information of intangible cultural heritage is divided into three levels for researching the metadata structure of corresponding levels. Then multimedia semantic model of intangible cultural heritage is constructed by the establishment of metadata mapping and integration relationship based on CIDOC CRM. Finally, the semantic modeling of multimedia for intangible cultural heritage is verified to provide a solution for the integration of heterogeneous multimedia data by taking the Dragon Boat Festival as an example.

**Keywords:** Intangible Cultural Heritage, Multimedia, Semantic Model, CIDOC CRM.

## 1 Introduction

Intangible cultural heritage (ICH) is the wisdom of human civilization of thousands of years, including the human emotions, the unspeakable meaning, and inestimable value. In recent years, traditional cultural ecology has been suffered varying degrees of damage with the rapid economic development, for example, some ICH has been extinct and some are on the verge of extinction [1].

The current protection of ICH mainly leans to digital preservation and visual reconstruction technology. The research of digital preservation technology concentrates in text, images, audio, video, which are record-based digital technology, and 3D models, motion, which are feature-based digital technology. Nevertheless, semantic metadata model of digitized multimedia resources is rarely involved in these researches. Intangible cultural heritage has more relationship with the knowledge level that includes accumulated, event-centric and non-material carrier experiences in the long term [2]. It is difficult to completely represent the complex relationship among different knowledge in the form of text, graphics, images, video, 3D model,

motion or other multimedia types. There are large numbers of heterogeneous knowledge in ICH, so it is particularly important to design an architecture and implementation techniques to combine and organize the heterogeneous multimedia data of ICH.

In this paper, the metadata structure for ICH is studied based on the analysis of multimedia data type hierarchy. And the relationship among multimedia metadata is extracted to describe the entities (such as concepts, objects, and events) of ICH and their complex semantic relationships. Then multimedia semantic model for ICH is constructed after the establishment of metadata mapping and integration relationship. Finally, the semantic modeling of multimedia for ICH is verified to provide a solution for the integration of heterogeneous multimedia data by taking the Dragon Boat Festival as an example.

## 2 The Design of Multimedia Metadata Framework for ICH

### Multimedia Data Hierarchy of ICH

Multimedia data of ICH is divided into three levels according to the degree of digital (Fig.1). At the bottom, there are record-based text and Grid-based multimedia data, such as text, images, audio, and video, etc. In the middle level, there are 3D models, motion, which are feature-based vector multimedia data. And semantic-based knowledge data is on the top level.

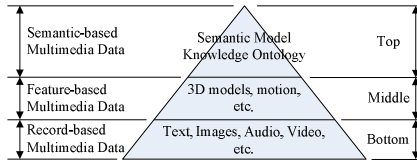


Fig. 1. Multimedia data contents hierarchy

Record-based multimedia data generally refers to basically digitizing multimedia data of ICH through text descriptions, audio or photographic cameras, etc. Feature-based multimedia data, which are abstracted to 3D models and motion, refers to the feature extraction and comprehensive collection of the abstract nature of ICH resources through the integrated use of 3D scanning and motion capture techniques. Those different levels of ICH multimedia data above are both separated digital data, and the inner links and differences among different data cannot distinguish by a simple metadata description, so those data could not reflect the connection of ICH knowledge.

The common concepts and semantic relations among concepts of the middle and bottom data are defined at different levels by semantic-based multimedia data, and then a model of semantic relations is constructed to clearly express semantic relations between multimedia data and ICH knowledge.

## The Design of Multimedia Metadata Structure for ICH

The mainstream metadata standards can be referred to design the ICH multimedia metadata framework in order to exchange and compatible with different standards, but the existing mainstream standard are too concise to describe the contents of data. ICH multimedia metadata, however, have features of complex Types, diverse formats, and all-inclusive contents. In order to achieve the full efficiency of ICH multimedia data management, mainstream metadata standards need to be extended necessarily to reflect different aspects of the individual needs of metadata by dividing the ICH contents into three levels from high to low. In this paper, ICH multimedia metadata structure model is mainly divided into three levels as shown in Fig.2.

The first layer is the core metadata of ICH resources that extracts common attributes of resources to achieve the unified management of various resources. The Core metadata contents, which is the basis for the development of follow-up metadata, is reference to the Dublin Core (DC) metadata standard [3], including 15 elements of Title, Subject, Description, Language, Source, Relation, Coverage, Creator, Publisher, Contributor, Rights, Date, Type, Identifier, and Format.

The second layer consists of metadata of 3D model, motion, video, audio, image and text. Since each media has unique properties compared to other media, the purpose of this additional layer is the embodiment of the unique attributes of each media.

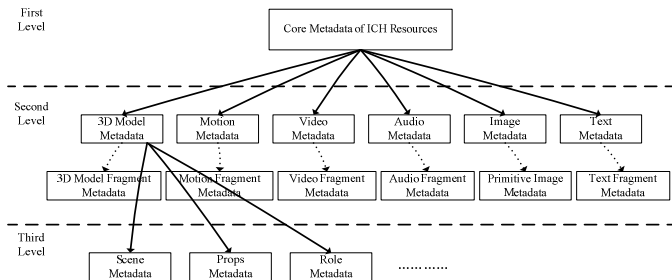


Fig. 2. Multimedia metadata framework for ICH

The third layer is the specific types of metadata developed depending on the category of specific types of ICH resources and according to different areas of expertise, categories of needs and applications. This part of the metadata, which is an extension of the upper metadata, is designed by combining with specific resource types, professional characteristics and needs, and under the participation of various research organizations.

## Multimedia Semantic Model for ICH

Metadata, which is used to identify and describe resources, does not express rich semantics and cannot reflect the complex relationship between multimedia data and knowledge of ICH.

The relationship among heterogeneous multimedia metadata is extracted to describe the entities (such as concepts, objects, and events) [4] and their complex semantic relationships through applying semantic model to the field of ICH. Semantic



model has a good concept hierarchy and support for logical reasoning, and can solve the heterogeneity of different multimedia data.

Semantic relations and relationship description of elements are dominantly declared by Concept - Relation - Concept statement in the semantic model. Multimedia semantic model for ICH is constructed by the establishment of metadata mapping and integration relationship.

### **3 Multimedia Semantic Model for ICH Based on CIDOC CRM**

#### **Semantic Reference Model for ICH**

In the field of cultural heritage, many ontology modeling methods are developed by some international organizations, including CDWA [5], AAT [6], TGN [7], ULAN [8], ABC Ontology [9], and CIDOC CRM [10], and so on.

CIDOC Conceptual Reference Model (CRM) is a core ontology and ISO standard (ISO 21127) and the overall scope of the CRM can be summarized in simple terms as the curated knowledge of scientific documents and cultural heritage [11]. CIDOC CRM contains various types of data objects, and aims to the depth and semantic relations of the multimedia data description. 90 entities and 148 properties that defined around the field of cultural heritage cover the knowledge of time, actor, historical facts, and humanity art, and so on. CIDOC CRM lays stress on the global description around an event, and connects related people, activity, and physical object together to represent the entities and properties according to their semantic relations. Therefore, it is appropriate to use CIDOC CRM to construct semantic models for ICH.

#### **Metadata Integration of ICH Based on CIDOC CRM**

Multimedia metadata structure of ICH is established based on the expansion of DC metadata standard, so metadata integration method for ICH based on CIDOC CRM is studied by taking DC metadata standard as an example in this paper. DC metadata standard includes 15 elements, which are Title, Subject, Description, Language, Source, Relation, Coverage, Creator, Publisher, Contributor, Rights, Date, Type, Identifier, and Format. CIDOC CRM defines 90 entities and 148 properties that can map and integrate all the core elements of DC.

Video data with a representative and broad will be taken as an example to describe the integration between CIDOC CRM and DC, because video contains both the image information, some audio information as well the characteristics of moving image.

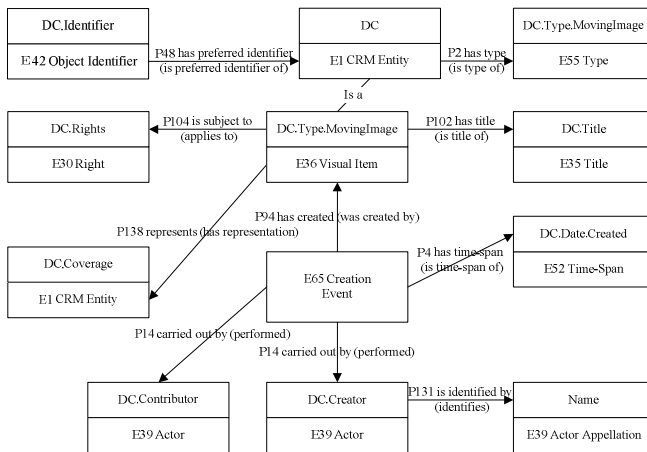
The type of resource object description (i.e., video) is defined by DC.Type in the description of video resources, and DC.Type can be mapped to the CRM Entity "E36 Visual Item", which includes the image identification mark and knowledge of images. Tab.1 lists the mapping relations between DC and CIDOC CRM for video resources descriptions by using the same ideas and methods above.

**Table 1.** Mapping relations Between DC and CIDOC CRM for video resources

DC	CIDOC CRM
Title	E35 Title
Subject	E1 CRM Entity
Description	E62 String
Source	E56 Language
Relation	E24 Physical Man-Made Stuff
Coverage	E81 Transformation
Creator	E36 Visual Item
Publisher	E39 Actor
Contributor	E82 Actor Appellation
Rights	E39 Actor
Date	E30 Right
Type	E52 Time-Span
Identifier	E36 Visual Item
Format	E42 Object Identifier

**Multimedia Semantic Model Construction for ICH Based on CIDOC CRM**

Semantic relations and relationship description of elements implicit in the definition of DC metadata, so it is difficult for computer to understand and handle them, otherwise, this relationship is explicit stated by Entities-Properties-Entities in CIDOC CRM. CIDOC CRM focuses on the description of the events around Entities. Therefore, the semantics of DC elements can be expressed by an event or a group of relationship in CIDOC CRM. For example, Entity “E12 Production” and “E65 Creation” in CIDOC CRM can be used to centralize all information related to the incident creation.



**Fig. 3.** Example of semantic model for ICH multimedia based on CIDOC CRM

The properties of different entities determined in CIDOC CRM are established by controlling the Entities-Properties-Entities relationship after the extraction and definition of entities. Entity "E36 Visual Item" and Entity "E30 Right" has a property of "P104 is subject to (applies to)", and the Entity "E30 Right" are just corresponding to DC. Rights. Therefore, the mapping between DC and CIDOC CRM is implemented not only by integrating related entities and classes, but also by defining the properties among entities to represent the semantic relations. Consequently, semantic model for ICH multimedia has been constructed as Fig.3 shows.

## 4 Applications

The Dragon Boat Festival, one of three major traditional Chinese festivals, has rich knowledge, and was designated as a World Intangible Heritage in Sep.30, 2009. A representative of a "Qu Yuan Temple Ritual" video resource, whose main content is a ritual with local characteristics held at the Qu Yuan Temple during the Dragon Boat Festival, is taken as an example to elaborate mapping relations between CIDOC CRM entities and DC metadata (Tab.2). And the multimedia semantic model as shown in Fig.4 is constructed based on properties of entities defined by CIDOC CRM.

**Table 2.** Mapping relations between DC and CIDOC CRM of "Qu Yuan Temple Ritual" video resource

DC	Cotents	CIDOC CRM
Title	Qu Yuan Temple Ritual	E35 Title
Subject	Dragon Boat Festival, Qu Yuan Temple, Ritual	E1 CRM Entity
Description	A ritual with local characteristics held at the Qu Yuan Temple during the Dragon Boat Festival	E62 String
Language	Chinese	E56 Language
Source	The real Qu Yuan Temple scene shooting at Zigui County, Hubei Province	E24 Physical Man-Made Stuff
Relation	A video set including folk activities of Dragon Boat Festival	E81 Transformation
Coverage	Inside and outside of Qu Yuan Temple during the Dragon Boat Festival at Zigui County, Hubei Province	E36 Visual Item
Creator	Wuhan Digital Media Engineering Technology Co.Ltd.	E39 Actor
Publisher	Wuhan Digital Media Engineering Technology Co.Ltd.	E82 Actor Appellation
Contributor	No	E39 Actor
Rights	Wuhan Digital Media Engineering Technology Co.Ltd.	E30 Right
Date	2011-03-09	E52 Time-Span
Type	Video	E36 Visual Item
Identifier	03250110309029	E42 Object Identifier
Format	.avi	E55 Type

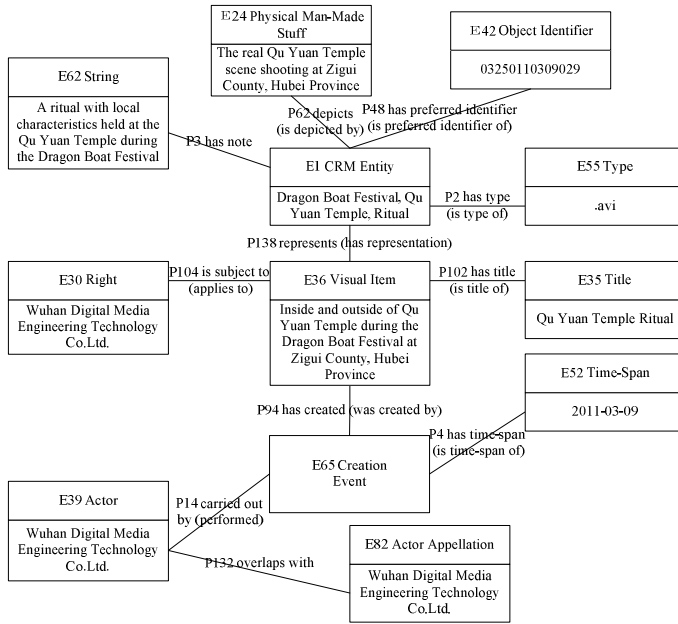


Fig. 4. Semantic model of "Qu Yuan Temple Ritual" video resource

## 5 Conclusion

Digital information of ICH is divided into three levels in order to achieve a comprehensive and efficient management for ICH multimedia data and the corresponding level metadata structure is studied by expanding mainstream metadata standard. Metadata mapping and integration relationship of ICH metadata are established based on CIDOC CRM, and the relationship among multimedia metadata is extracted to describe the entities (such as concepts, objects, and events) of ICH and their complex semantic relationships. Then multimedia semantic model for ICH with a good concept hierarchy and support of logical reasoning and solve the heterogeneity of different multimedia data of ICH, is constructed. Finally, the semantic modeling of multimedia for intangible cultural heritage is verified to provide a solution for the integration of heterogeneous multimedia data by taking the Dragon Boat Festival as an example.

**Acknowledgments.** This paper is financially supported by the Key Cooperation Project in Industry, Education and Research of Guangdong Province and Ministry of Education (No. 2009A090100023), and the Cooperation Project in Industry, Education and Research of Guangdong Province and Ministry of Education (No.2009B090200057), and National Natural Science Foundation of China (No.41101406).

## References

1. Long, X.-Q.: Connotation and Characters of Intangible Cultural Heritage and Its Protection. *Journal of Hubei Institute for Nationalities (Philosophy and Social Sciences)* 249(5), 47–52 (2006)
2. Tan, G., Hao, T., Zhong, Z.: A Knowledge Modeling Framework for Intangible Cultural Heritage Based on Ontology. In: *Proceedings of the 2nd International Symposium on Knowledge Acquisition and Modeling (KAM 2009)*, vol. 1(11), pp. 304–307 (2009)
3. Dublin core metadata initiative (DB/OL) (2000), <http://dublincore.org>
4. Sinclair, P., Addis, M., Choi, F., et al.: The use of CRM core in multimedia annotation. In: *The Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006)*, pp. 22–26. CEUR-WS, Tilburg (2006)
5. Categories for the Description of Works of Art (CDWA) (EB/OL) (2011), [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/](http://www.getty.edu/research/publications/electronic_publications/cdwa/)
6. Art and Architecture Thesaurus (EB/OL) (2011), <http://www.getty.edu/research/tools/vocabularies/aat/index.html>
7. Thesaurus of Geographic Names (EB/OL) (2011), <http://www.getty.edu/research/tools/vocabularies/tgn/>
8. Union List of Artist Names (EB/OL) (2011), <http://www.getty.edu/research/tools/vocabularies/ulan/>
9. Lagoze, C., Huster, J., The, A.B.C.: *Ontology and Model* (EB/OL) (2009), [http://metadata.net/harmony/JODI\\_Final.pdf](http://metadata.net/harmony/JODI_Final.pdf)
10. Crofts, N., Doerr, M., Gill, T., et al.: Definition of the CIDOC Conceptual Reference Model (Version 5.0.2) (S/OL) (2010), [http://www.cidoc-crm.org/official\\_release\\_cidoc.html](http://www.cidoc-crm.org/official_release_cidoc.html)
11. Doerr, M., Ore, C., Stead, S.: The CIDOC Conceptual Reference Model – a New Standard for Knowledge Sharing. In: *Proceedings of the 26th International Conference on Conceptual Modeling (ER 2007)*, pp. 51–56. Springer, London (2007)

# Improved Photonic Crystal 90° Bends for Millimeter—Wave Transmission

Q.Z. Xue<sup>1</sup> and S.J. Wang<sup>2</sup>

<sup>1</sup> Key Laboratory of High Power Microwave Sources and Technologies, Institute of Electronics, Chinese Academy of Science, Beijing, 100190, China

<sup>2</sup> Key Laboratory of High Power Microwave Sources and Technologies, Institute of Electronics, Chinese Academy of Science, Beijing, 100190, China  
Graduate University, Chinese Academy of Sciences, Beijing 100190, China  
Qianzhong\_xue@mail.ie.ac.cn

**Abstract.** Metallic waveguides are popular used in microwave and millimeter wave transmission, but they are not easy to bend. 2D metallic photonic crystal structures have broad band gap and can be bended easily. In this paper improved photonic crystal 90° bends with high transmission and broad bandwidth covering over 220GHz atmosphere windows are designed by optimizing structure on the bending corner, and the transmission spectrum of the bends are given.

**Keywords:** Metallic waveguide, metallic photonic crystal, bend waveguide.

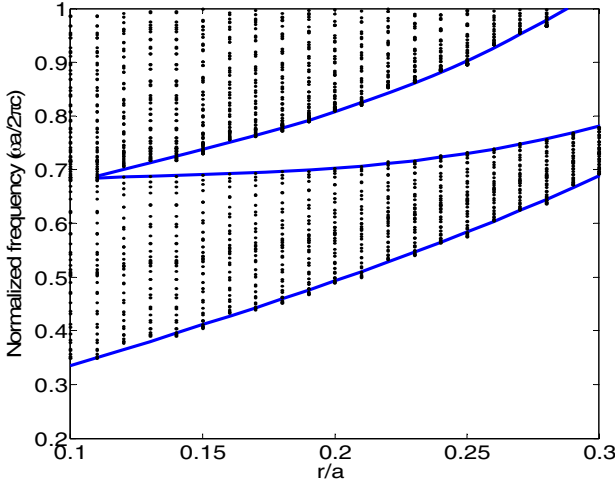
## 1 Introduction

Metallic waveguides are popular used in microwave and millimeter wave transmission, because of their low propagation loss. Waveguides have to have an ability of bending for interconnection to other devices. Traditional metallic waveguides are not easy to bend and bends will introduce losses caused by reflection at the corners, in order to prevent this, the radius of bends will become larger, that is not suitable for optical integrated circuit. Photonic crystal (PC) with 2D metallic rods is known as a waveguide without bending loss. [1] Electromagnetic waves whose frequencies are in the photonic band gap (PBG) cannot propagate in the photonic crystal structure, so no power will be radiated out of the structure as electromagnetic waves around the bend. We can improve the 90° bend waveguides by reduce the reflection caused by bend.

## 2 The Band Gap of TM Mode in Perfect Photonic Crystal

Fig.1 shows Band gap map of TM mode in square lattice metallic photonic crystal,  $a$  being the lattice constant,  $r$  being the radius of metallic rods. Metallic photonic crystals (MPC) have important advantages over dielectric photonic crystal (DPC) [2, 3]. The band gaps of MPC TM modes are much wider than that of DPC TM modes,

and there exists a wide zero order band gap of the modes, which is a cutoff analogous to that in a conventional waveguide and exists for all values of  $r/a$ , whereas there is no such cutoff in dielectric lattices for either TE or TM modes. In the other hand, the width of the global TM band gap in the metallic lattice increases with increasing  $r/a$ , whereas the global TE and TM band gaps in dielectric lattices typically close as the ratio  $r/a$  increases. These properties equip MPC with some advantages that DPC hasn't, that is, wide band gap, which is convenient for mode selecting and enhancing the stability of the operating mode. And if an operating mode in the zero order band gap is selected, the size and weight of devices can be minimized so as for integration.



**Fig. 1.** Band gap map of TM mode in square lattice metallic photonic crystal,  $a$  being the lattice constant,  $r$  being the radius of metallic roads

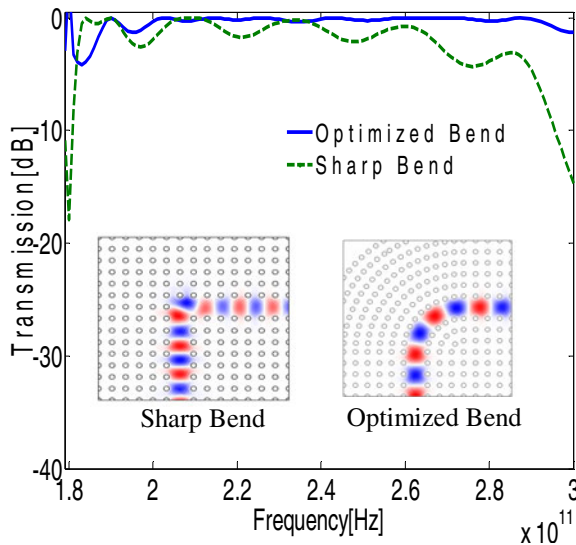
### 3 Design

The considered photonic crystal waveguide model consists of copper rods in air on a square array. Making sure the waveguide can guides millimeter wave at the center frequency 220GHz, we choose lattice constant  $a=0.52\text{mm}$ , radius of copper roads  $r=0.2a$ . Fig.2 shows the transmission characteristics of the sharp  $90^\circ$  bend waveguide. It is found that transmission fluctuates between  $-3\text{dB}$  and  $0\text{ dB}$  at the range of frequency from 183GHz to 270GHz and doesn't reach design demand. Propagation loss mainly is caused by back reflection. The reflection coefficient can be given by[4]

$$R(\omega) = \left[ 1 + \left( \frac{2k_1(\omega)k_2(\omega)}{[k_1^2(\omega) - k_2^2(\omega)]\sin[k_2(\omega)L]} \right)^2 \right]^{-1} \tag{1}$$

It is seen from Eq.(1) that when  $k_1^2(\omega) - k_2^2(\omega)$  tends to 0 , reflection coefficient can be eliminated. Here,  $k_1$  is the wave vector in straight channel,  $k_2$  is the wave vector in bend channel,  $L$  the length of the bend.

In order to improve transmission characteristics of the bend, the curve structure is designed by placing the bending corner with 90° arc of rods and increasing the number of rods on the bending arcs without varying the size of rods. For a comparison, the transmission of the optimized bend waveguide and the corresponding structure are also shown in Fig.2.



**Fig. 2.** Transmission spectrum of 90°sharp bend and optimized bend MPC waveguide. Insets show the corresponding bend structures and Ez field pattern in the vicinity of the bends for frequency f= 220GHz

It is found from Fig.2 that the transmission of the optimized bend waveguide is improved greatly. 0.5-dB transmission bandwidth for the optimized bend waveguide is 93GHz between  $2.0 \times 10^{11}$  and  $2.93 \times 10^{11}$ Hz, While 3-dB transmission bandwidth for the sharp band is 90GHz.

### 4 Conclusion

Transmission characteristics of the 90°bend waveguide is studied. The 90°bend waveguide with high transmission is designed based on curved bend structure . The simulation results show that high transmission with 0.5-dB frequency band of 93GHz near frequency 220GHz are reached.



**Acknowledgments.** This work is supported by the National Natural Science Foundation of China under Grant 60871051.

## References

1. Kokubo, Y., Kawai, T., Ohta, I.: Photonic Crystal Waveguides with Inner Metal-Coated Posts at Millimeter-Wave Frequency Range. In: 34th European Microwave Conference, Amsterdam, pp. 1137–1140 (2004)
2. Li, Y.-L., Xue, Q.-Z., Du, C.-H., Hao, B.-L.: Modified Finite-Difference Frequency-Domain Method for Two-Dimensional Metallic Photonic Crystal Analysis. *Acta Physica Sinica*, 2556–2563 (2010)
3. Degirmenci, E., Surre, F., Landais, P.: Improved Photonic Crystal Based 90° Bends For THz Transmission, pp. 754–755. IEEE (2010)
4. Mekis, A., Chen, J.C., Kurland, I., Fan, S., Villeneuve, P.R., Joannopoulos, J.D.: High Transmission through Sharp Bends in Photonic Crystal Waveguides. *Physical Review Letters*, 3787–3790 (1996)

# Simulation of Algorithm of Binary Offset Carrier Signal

Xu Jia and Bi Mingxue

Beijing Institute of Technology, Beijing, China  
{xujia\_virginia,bmx418}@163.com

**Abstract.** BOC (Binary Offset Carrier) modulation accommodates to the new requirement of satellite navigation development. It can improve the precision of the measurement and also possess the higher ability of anti-interference. This paper covers the studies on the theory and characteristic of BOC signal, on the frequency spectrum, power spectrum and autocorrelation of BOC modulation, on the principle of BOC acquisition, and on the principle of BOC tracking from the sight of carrier tracking and code tracking separately. The simulation results are shown such as BOC signal based on BOC modulation, the acquisition of BOC signal based on the method of parallel code phase search acquisition, both carrier tracking loop and bump-jump code tracking loop, and acquisition signal tracking based on the above two loops. Finally organizing and analyzing the acquired statistics are given.

**Keywords:** BOC, auto-correlation function, tracking.

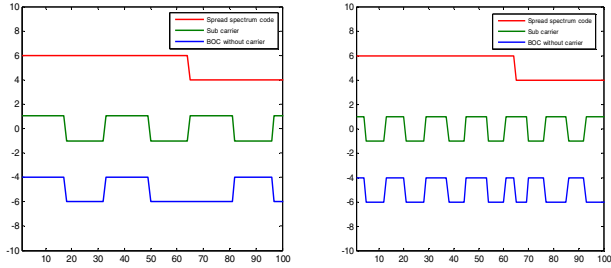
## 1 Introduction

Nowadays, GPS utilizes BPSK as its main modulation model. However, the power of signals modulated by BPSK focus on the carrier frequency, which suffers from the problem of lower transmission efficiency, deficiency of anti-interference, etc. Due to the limitation in the aspects of techniques and regulations as well as the requirement of satellite navigation, researchers put forward another modulation method -- binary offset carrier [1, 2]. The power spectrum density of BOC is made of main lobes and side lobes, and correlation function has several peak values. Its unique splitting characteristic of power spectrum achieves the co-utilization of frequency band and alleviates the interference among signals [3]. Moreover, its correlation function is much sharper than that of BPSK with the same bit rate, and it has a higher acquisition precision. Therefore, BOC is useful for the satellite navigation in the aspects of calculation of delay and ranging.

## 2 Characteristics of BOC Signal

Basically, BOC(m,n) or BOC( $mf_0, nf_0$ ) signal is created by modulating a sine wave carrier with the product of a PRN spreading code and a square wave subcarrier with binary  $\pm 1$  values. Here,  $f_0=1.023$  MHz, subcarrier frequency  $f_s = mf_0$  in MHz, and spreading code rate  $f_c = nf_0$  Mchip/s [5, 6].

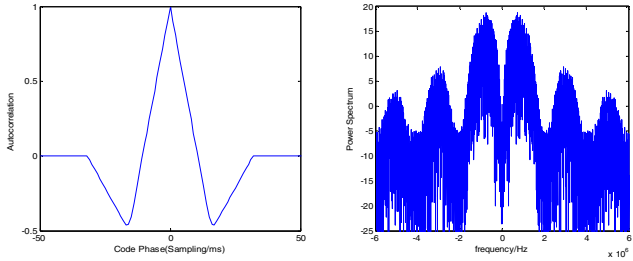
Fig.1 simulates the time domain wave form of different BOC signals. They are spread spectrum code, sub-carrier and BOC signal without carrier separately. Additionally, Fig.2 is the ACF and power spectrum of several different BOC signals.



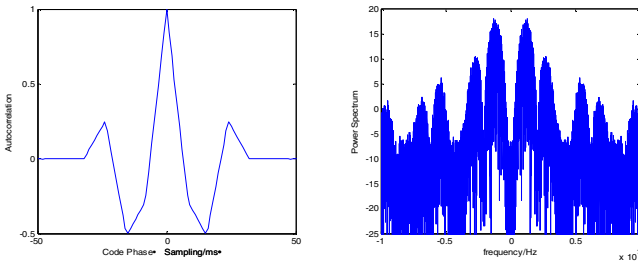
(a) BOCsin(1,1)

(b) BOCcos(2,1)

**Fig. 1.** Different BOC signals in time domain



**Fig. 2(a).** BOCsin(1,1) ACF and Power Spectrum



**Fig. 2(b).** BOCcos(1,1) ACF and Power Spectrum

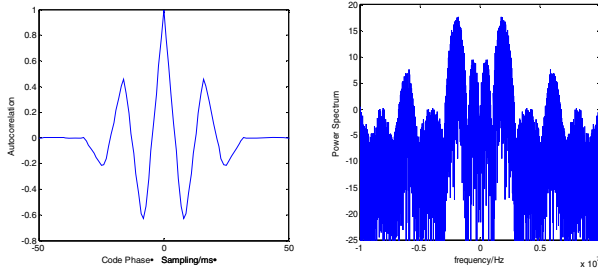


Fig. 2(c). BOCsin(2,1) ACF and Power Spectrum

### 3 BOC Acquisition

The purpose of acquisition is to get coarse values of carrier frequency and code phase of the satellite signals. Generally, the amount of search steps in the code phase dimension (0 to 1022 chips) is significantly larger than that of the frequency dimension ( $\pm 10$  kHz in steps of 500 Hz). The method parallelized the frequency space search eliminating the necessity of searching through the 41 possible frequencies, requiring 1023 steps. If the acquisition could be parallelized in the code phase dimension, only 41 steps should be performed in the parallel code phase search acquisition algorithm.

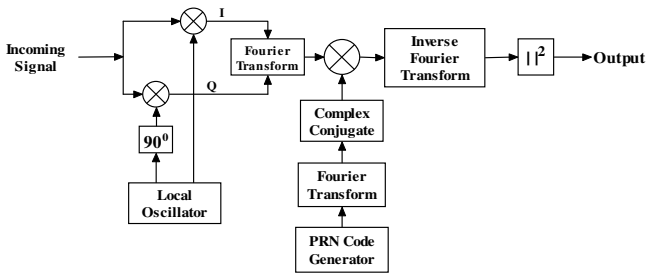


Fig. 3. Parallel Code Phase Search Algorithm

The incoming signal is multiplied by a locally generated carrier signal. Multiplication with the signal generates I signal, and multiplication with a  $90^\circ$  phase-shifted version of the signal generates the Q signal. The I and Q signals are combined to form a complex input signal  $x(n) = I(n) + jQ(n)$  to the DFT function. The generated PRN code is transformed into the frequency domain and the result is complex conjugated. The Fourier Transform of the input is multiplied with the Fourier Transform of the PRN code. The result of the multiplication is transformed into the time domain by an inverse Fourier Transform. The absolute value of the output of the inverse Fourier Transform represents the correlation between the input and the PRN code. If a peak is present in the correlation, the index of this peak marks the PRN code phase of the incoming signal.

After correlation of I and Q, the output in this step is

$$\begin{aligned}
 U_{ki}(l) &= I_{ki} - jQ_{ki} = \sum_{n=0}^{N-1} s_n(t_n) S_{ref}^*(t_n; \hat{\tau}_k, \hat{f}_{dl}) \\
 &= \sum_{n=0}^{N-1} (s_n(t_n) e^{-j2\pi\hat{f}_{dl}t_n}) G_{T_c}^{ref}(t_n - \hat{\tau}_k)
 \end{aligned}
 \tag{1}$$

If

$$s_{nb}[n] = s_n(t_n) e^{-j2\pi\hat{f}_{dl}t_n}, \quad G_{T_c}^{ref}[n-k] = G_{T_c}^{ref}(t_n - \hat{\tau}_k) \tag{2}$$

Then

$$U_{ki}(l) = \sum_{n=0}^{N-1} (s_{nb}[n] G_{T_c}^{ref}[n-k]) = s_{nb}[k] \otimes G_{T_c}^{ref}[-k] \tag{3}$$

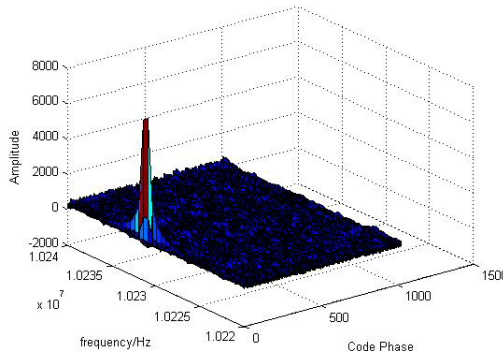
Then

$$\begin{aligned}
 DFT\{U_{ki}(l)\} &= DFT\{s_{nb}[k] \otimes G_{T_c}^{ref}[-k]\} \\
 &= DFT\{s_{nb}[k]\} \cdot DFT\{G_{T_c}^{ref}[-k]\} \\
 &= DFT\{s_{nb}[k]\} \cdot (DFT\{G_{T_c}^{ref}[k]\})^*
 \end{aligned}
 \tag{4}$$

So we get

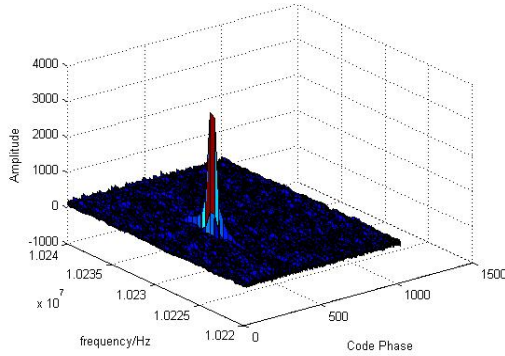
$$\begin{aligned}
 U_{ki}(l) &= IDFT\{DFT\{U_{ki}(l)\}\} \\
 &= IDFT\{\{s_{nb}[k]\} \cdot (DFT\{G_{T_c}^{ref}[k]\})^*\}
 \end{aligned}
 \tag{5}$$

The length of pseudo-code is 1023×60ms, sampling frequency is  $f_{\text{sampling}}=32f_0=32 \times 1.023\text{MHz}$ , carrier frequency is  $f_{\text{carrier}}=10f_0=10 \times 1.023\text{MHz}$ , subcarrier frequency of square wave is  $f_{\text{subcarrier}}=10f_0=1.023\text{MHz}$ . Simulate the acquisition of BOC(1,1), phase of subcarrier is  $\sin(\cdot)$ . In this simulation, as to the sampling frequency  $f_{\text{sampling}}=32f_0$ , seven chips correspond to the  $7 \times 32$  sampling points, 300 chips to the  $300 \times 32$  sampling points. From the results, we could get the highest value of ACF, in which the corresponding delay and Doppler shift will be regarded as the estimation values of the delay and shift. Doppler shift is 1234Hz, chip delay is 7 chips, the SNR is 0dB, amplitude of modulation signal is 1.0, 0.5, 0.2. The 3D results of acquisition are shown in Fig.4.

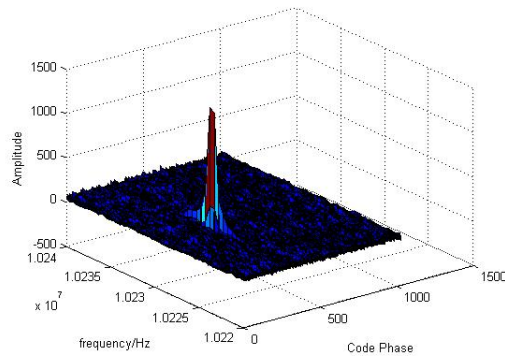


(a) No noise, fdoppler=1234Hz, chip delay = 7chips, amplitude = 1.0

**Fig. 4.** The 3D Results of Acquisition



(b) No noise, fdoppler= -1234Hz, chip delay = 300chips, amplitude = 0.5



(c) No noise, fdoppler= -1234Hz, chip delay = 300chips, amplitude = 0.2

**Fig. 4.** (Continued)

**Table 1.** Simulation Parameters I

No	SNR (dB)	Ratio of Peak Value to First Side Lobe	Carrier Frequency(Hz)	Code Phase of acquisition	amplitude of modulation signal
a	No noise	14.125892	10231233.170	7	1.0
b	No noise	13.336927	10228766.830	300	0.5
c	No noise	13.336927	10228766.830	300	0.2

Comparing the estimated values and the results of the acquisition, we can deem it as a successful acquisition. The setting of acquisition threshold is based on the ratio of peak values and the first side-lope. In the above table, differences are small. Therefore, this data is not very related to the amplitude of signal. In the meantime, the difference of the carrier frequency of acquisition is around  $\pm 1$ Hz.

Additional Simulation: Adding noise to the incoming signal to verify the availability of this acquisition method.

**Table 2.** Simulation Parameters II

No	SNR (dB)	Ratio of Peak Value to First Side Lobe	Carrier Frequency(Hz)	Code Phase of acquisition	amplitude of modulation signal
a	-20	3.425417	10231233.167	7	1.0
b	-20	3.120208	10228766.830	300	0.5
c	-20	3.545765	10228766.830	300	0.2

**Table 3.** Simulation Parameters III

No	SNR (dB)	Ratio of Peak Value to First Side Lobe	Carrier Frequency(Hz)	Code Phase of acquisition	amplitude of modulation signal
a	-25	2.363926	10231233.170	7	1.0
b	-25	2.121877	10228766.830	300	0.5
c	-25	1.931420	10228766.830	300	0.2

**Table 4.** Comparison of ratio of Peak value to first side lobe

SNR(dB)	$f_{doppler}=1234\text{Hz}; \text{chip delay}=7\text{chips}, \text{amplitude}=1$	$f_{doppler}=-1234\text{Hz}; \text{chip delay}=300\text{chips}, \text{amplitude}=0.5$	$f_{doppler}=-1234\text{Hz}; \text{chip delay}=300\text{chips}, \text{amplitude}=0.2$
No noise	14.125892	13.336927	13.336927
-20	3.425417	3.120208	3.545765
-25	2.363926	13.336927	1.931420

With the decreasing of SNR, the peak value of acquisition also reduces.

## 4 BOC Tracking

The main purpose of tracking is to refine the values got from acquisition, which provides only rough estimates of the frequency and code phase parameters. A basic demodulation scheme is shown in Fig.5.

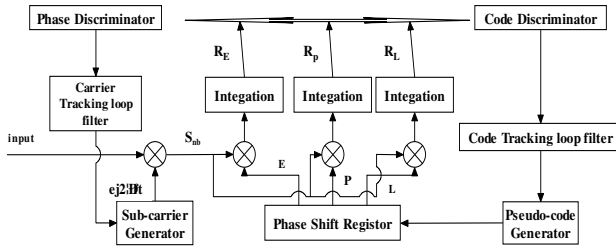


Fig. 5. Basic Demodulation Scheme

The main function of carrier tracking is to demodulate the output signal from code tracking loop, then to get the navigation datagram and the calculation of Doppler-Shift for the measuring accuracy of receiver. To demodulate datagram successfully, it is necessary to produce an accurate carrier signal.

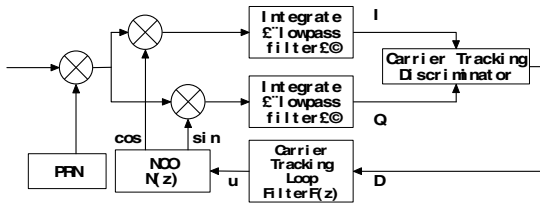


Fig. 6. Carrier Tracking Loop

The signal of I arm is:

$$D^k(n) \cos(\omega_{IF}n) \cos(\omega_{IF}n + \varphi) = \frac{1}{2} D^k(n) \cos(\varphi) + \frac{1}{2} D^k(n) \cos(2\omega_{IF}n + \varphi) \tag{6}$$

The signal of Q arm is:

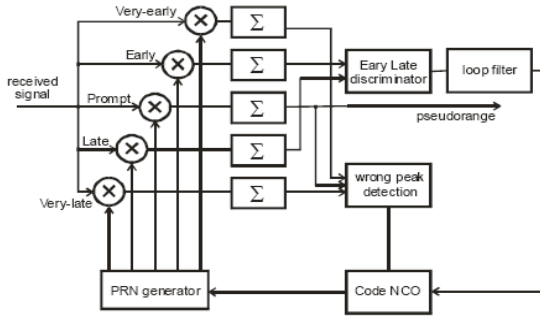
$$D^k(n) \cos(\omega_{IF}n) \sin(\omega_{IF}n + \varphi) = \frac{1}{2} D^k(n) \sin(\varphi) + \frac{1}{2} D^k(n) \sin(2\omega_{IF}n + \varphi) \tag{7}$$

$\varphi$  is the difference between the phase of input signal and the local carrier. After lowpass filter, the following two signals remain:  $I^k = \frac{1}{2} D^k(n) \cos(\varphi)$   $Q^k = \frac{1}{2} D^k(n) \sin(\varphi)$ , in which  $D^k(n)$  is the Costas discriminator.

The goal for a code tracking loop is to keep track of the code phase of a specific code in the signal. The output of such a code tracking loop is a perfectly aligned replica of the code. The way to handle the problem with tracking of correct peak of BOC modulation correlation function is the technique -- very early -- very late (VEVL) correlator, also known as “bump-jump” method. In comparison to the classical early – late correlator structure, VEVL has a further couple of early and late taps. This extra couple of taps is adjusted to track the side-peaks of correlation function. The early and late taps together with prompt tap are intended for tracking

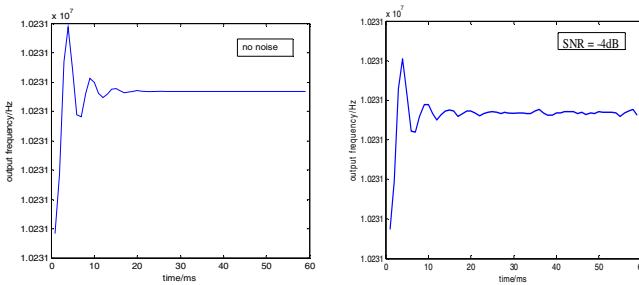


the correct (centre) peak of the correlation function like in the classical early – late correlator. The spacing (a correlator width) is adjusted to enable tracking the narrow peak of particular type of BOC correlation function. The additional very early – very late taps are set to watch the side-peaks of the correlation function. When the correlator tracks the correct correlation peak, the prompt tap output is greater than those from very-early and very-late taps. In case of repetitively greater output from very-early or very-late taps, the wrong peak tracking is declared. Then the phase of a local signal replica is adjusted to restore the tracking of the correct peak again.

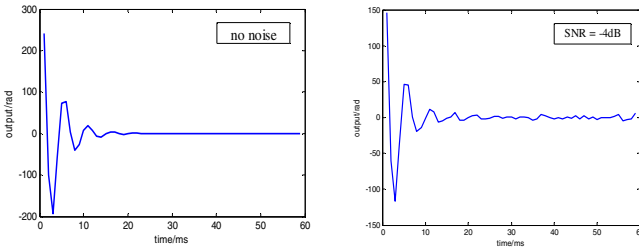


**Fig. 6.** Structure of VEVL Correlator

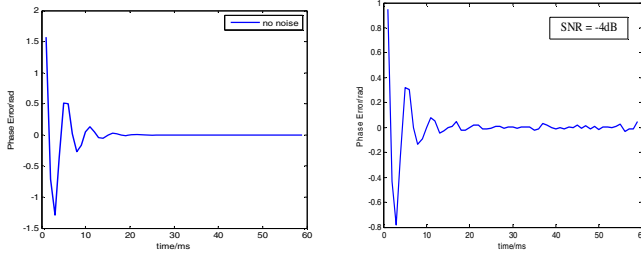
Fig.7 to 12 is the output results of each block in the tracking. (Left Fig. has no noise, right Fig. has SNR with -4dB)



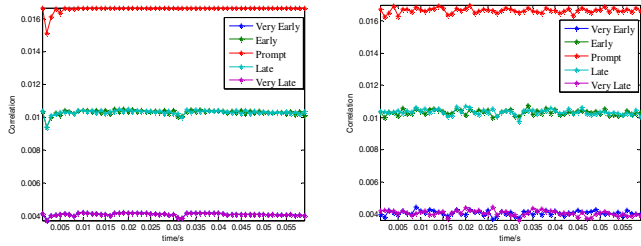
**Fig. 7.** Output of NCO Carrier Generator



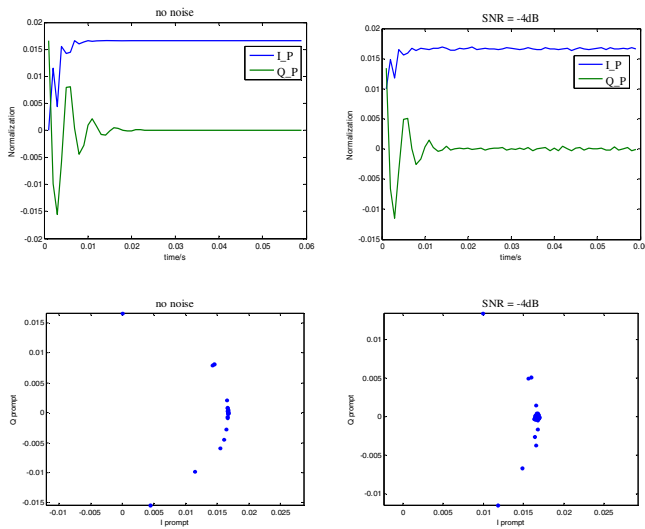
**Fig. 8.** Carrier Loop Filter



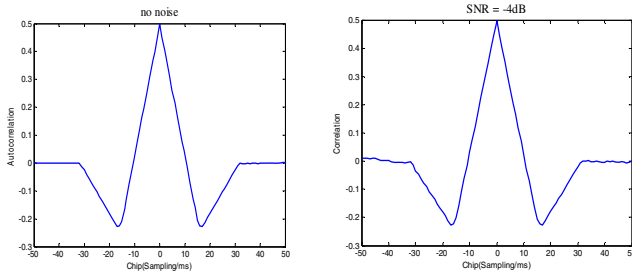
**Fig. 9.** Carrier Loop Discriminator



**Fig. 10.** Prompt, early, late, very-early and very-late correlation results



**Fig. 11.** Output of prompt code with in-phase and quadrature, eye-diagram



**Fig. 12.** ACF of BOC Subtracting Carrier

**Table 5.** Average Results of Tracking

SNR(dB)		No noise	-4
<b>Carrier tracking loop</b>	<b>NCO carrier frequency(Hz)</b>	10231234.023	10231234.094
	<b>Discriminator output (e-006)</b>	448.489	736.859
	<b>Filter output (e-003)</b>	-39.452	39.346
<b>Code tracking loop</b>	<b>Code phase</b>	7	7
	<b>Discriminator output (e-006)</b>	0.001815	-7.588
	<b>Filter output (e-003)</b>	0.000	0.000
<b>In-phase prompt (e-003)</b>		16.667	16.6999
<b>Quadrature prompt (e-006)</b>		7.475	12.199

**Table 6.** Average Results of Tracking II

SNR(dB)		4	-10
<b>Carrier tracking loop</b>	<b>NCO carrier frequency(Hz)</b>	10231234.013	10231234.057
	<b>Discriminator output (e-006)</b>	252.171	-313.749
	<b>Filter output (e-003)</b>	-29.245	-114.811
<b>Code tracking loop</b>	<b>Code phase</b>	7	7
	<b>Discriminator output (e-006)</b>	1.520	141.051
	<b>Filter output (e-003)</b>	0.000	0.243902
<b>In-phase prompt (e-003)</b>		16.659	16.6489
<b>Quadrature prompt (e-006)</b>		4.219	-6.749

When the tracking loop is locked, the output of carrier frequency fluctuates near 10.23MHz+1234Hz and is very small. Meanwhile, the outputs of carrier loop discriminator and carrier loop filter are close to zero. After tracking successfully, the output of prompt in-phase arm is larger than that of quadrature arm. The values of very-late and very-early arms are equal and smaller than those of late and early arms. The eye-diagrams illustrates that after tracking successfully, the dots are on the x-axis. In the case of failure, the dots are scattered near the x-axis. When the SNR is different, the output will fluctuate and the amplitude is different.

## 5 Conclusion

With the popularity of global satellite navigation, GPS has been updated to achieve high performance for civil use. It is valuable to do research about BOC which is a new type of modulation function. This paper focuses on the points as follows.

First, it mainly analyses modulation property of BOC and does some comparison with BPSK, which shows the differences between them clearly. The power spectrum and auto-correlation function of different BOC can be gotten.

Second, as to the acquisition of BOC, there are basically three types of acquisition methods, i.e., serial search acquisition, parallel frequency space search acquisition, parallel code phase search acquisition-briefly. Taking the parallel code phase search acquisition, I simulated the signal acquisition under different circumstances, such as different SNR.

Finally, in the part of tracking, there are mainly two parts-carrier tracking and code tracking. Different from the acquisition, which is a kind of rough estimation, tracking provides accurate estimation of frequency and delay for the received signal. In carrier tracking, the theory of carrier tracking loop is the main point. In code tracking, it is important to learn the classic structure of Early-Late tracking in BPSK. As to BOC, the bump-jump method overcomes the shortcoming of Early-Late tracking, which is called very early-very late (VEVL) correlator. In the simulation experiment, the tracking of BOC is achieved and the final data is got. Additionally, the correlator width could be changed according to different incoming signal and the currently tracking results, which should be explored further in the future.

## References

1. Kovar, P., Vejrazka, F., Seidl, L., Kacmarik, P.: Galileo Receiver Core Technologies. *Journal of Global Positioning Systems* 4, 176–183 (2005)
2. Borre, K., Akos, D.M., Bertelsen, N., Rinder, P., et al.: *A Software-Defined GPS and Galileo Receiver*. Springer, Heidelberg
3. Barker, B., Betz, J., Clark, J., Correia, J., Gillis, J., Lazar, S., Rehorn, K., Straton, J.: Overview of the GPS M Code Signal. MITRE Technical Papers Archive (2005) (cit. November 4, 2004)
4. Fante, R.: Unambiguous Tracker for GPS Binary-Offset-Carrier Signals. MITRE Technical Papers Archive (2009) (cit. November 4, 2004)
5. Fine, P., Wilson, W.: Tracking Algorithm for GPS Offset Carrier Signals. In: *Proceedings of ION National Technical Meeting, Institute of Navigation*, vol. 1, pp. 671–676 (2008)
6. Kovář, P., Vejražka, F., Seidl, L., Kačmařík, P.: Galileo receiver core technologies. *Journal of Global Positioning Systems* 4, 176–183 (2010)

# The Performance Test Research for Application System of Internet of Things

Song Qizhu, Ma Aiwen, Peng Xiao, Wang Junfeng, and Zhang Junchi

The State Radio\_monitoring\_center Testing Center, China  
{Songqzhu,maaiwen,pengxiao,wangjff,zhanjunchi}@srtc.org.cn

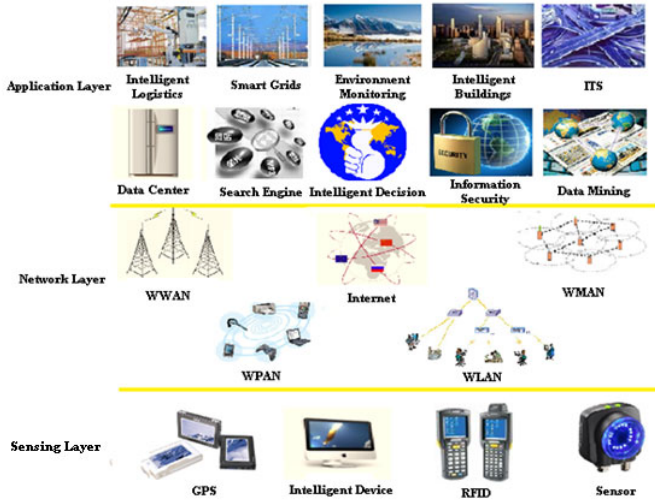
**Abstract.** With the development of application system of Internet of things, the testing requirements about the wireless sensing device are also on the rise. There are two testing requirements. One is the national requirements on spectrum management. In order to ensure the normal operation of the radio services, state must manage the various micro-power and short distance wireless devices. The other requirement is the users' needs on the system performance. The user often requests the authoritative third party to test the wireless devices in order to obtain the evidence for the system designing and the hardware selection. Basing on the performance testing of the application system, the paper researched the test contents and methods of wireless sensing devices on Internet of things, which can ensure the development of IOT industry.

**Keywords:** Internet of things, Performance test, RFID.

## 1 Introduction

IOT (Internet of Things) is the information industry revolution following the computer, Internet and the mobile communication. In August 2009, China proposed the strategy of sensation China. In March 2010, IOT was classified as one of the major national strategic industries. IOT was initially defined as that all things are connected to Internet by the sensing devices such as the radio frequency identification (RFID) devices and bar codes and so on. With the development of the technologies, the generalized concept of IOT is that on the basis of the information network, through the materials coding and tagging system, the things can be connected with webs by the data collection and sensing technology according to the standardized protocols, and thus it can realize the intelligent revolution about identification, positioning, tracking, monitoring and management of the things.

IOT includes the sensing layer, the network layer and the application layer, as shown in Figure 1.



**Fig. 1.** The structure of IOT

In the sensing layers, the data collection and perception is mainly used for collecting physical events and data, including various physical quantities, identifications, audio and video data. The data collection of IOT involves multimedia information collection, such as sensors, RFID, QR Code, real-time positioning technology and so on.

In the network layer, in order to transmit information accessibly, more reliably and more security, it needs integrate the sensor network, the mobile communication system and the Internet. It has rapidly developed for more than 10 years, and the mobile communication system, the Internet and other technologies have been much more mature, which can meet the needs of IOT data transmission.

In the application layer, it mainly involves the supporting platform layer and the application service layer. The supporting platform layer can support the information synergy, sharing and communication of multi-industry, multi-use and multi-system. The application service layer includes the intelligent transportation, the intelligent medical treatment, the intelligent house and home, the intelligent logistics, the intelligent electricity and other industry applications.

## 2 The Performance Test Contents

In IOT, RFID is the most representative technology. In the application of RFID, the users often request the authoritative third party to test the hardware device's performance of RFID system in the lab environment and the practical environment, in order to ensure the RFID system to meet the requirements of the project implementation.

According to the user's demand, the test objects usually contain readers and tags, and the test parameters include physical performance, air interface and transmission protocol, environment adaptability, electromagnetic compatibility, security reliability, the read and write distance and speed, etc. Relevant test items include the reader's test items and the tag test items, the reader's test items include the interface protocol, data rate, work temperature, work humidity, read and write distance, vibration, impact, collision, security, and electromagnetic immunity, etc. The tag's test items include air interface protocol, frequency, modulation, storage capacity, work environment, anti-radiation, alternating electromagnetic field resistance, impact, mechanical vibration, free drop and static electricity test, etc. These test items are tested under ideal electromagnetic environment. In order to ensure the system work properly, it is necessary to test the interoperability for readers and tags in the actual application environment. Based on these tests, it forms a testing chain which includes product test, system test and application test.

The product test faces to the level test of RFID devices, and includes function test, physical characteristics test, electromagnetic characteristics test and air interface test. The physical characteristics test is mainly about durability and anti-destroyed, such as distortion, vibration, friction, impact, sealing, X light irradiation and UV irradiation, etc. The electromagnetic characteristics test means the test for anti-electromagnetic performance, such as electrostatic discharge, surge, instantaneous pulse, alternating electric field and cross variable magnetic field, etc. The air interface test includes radiation power, frequency tolerance, occupied bandwidth, dwell time, radiated spurious, signal rise time, fall time, settling time, modulation depth, signal interaction time, transmission rate, and so on.

The system test includes system performance test in the ideal environment and EMC test. The system performance test items in the ideal environment include the distance, range, time, rate of the reading and writing, and the thing's moving speed. By this way to judge whether the tags support identification, data read, data write and multiple tags' anti-collision, and whether the readers could identify the read and write data on single label and multiple labels. EMC test is similar to other wireless devices' EMC test, it refers to test parameters and test methods in the international standards and the states standards.

The application test includes system performance test in the specific application environment and in complex electromagnetic environment. In specific application environment, system performance test should consider the environmental factors for the performance test, the factors include articles stockpiled, antenna location, read/write success rate, read/write rate, articles move speed, and so on. In complex electromagnetic environment, the system performance test need consider the factors of multiple wireless system coexistence. In this condition, it requires that the electromagnetic signal strength of all systems could be controlled.

### **3 The Test Environment**

In the perception layer of IOT, there are many wireless devices, while these wireless devices are not only electromagnetic interference source, but also can be interfered by

other wireless devices more or less. How to make the huge numbers of wireless devices work normally and not interfere each other, and how to plan and manage the wireless spectrum reasonably and effectively in IOT, many problems should be solved. In order to solve these problems and provide the basis for planning and wireless spectrum management and technology supporting for applications, it needs five test environments which are based on the test chain features, including device level physical characteristics test environment, device level EMC test environment, device and system benchmark test environment, air interface test environment and interoperability test environment.

The device level physical characteristics test environment is used for wireless sensing device test such as physical dimension, warping, vibration, impact, resistance to fall, seal, work temperature and humidity, resistance to corrosion and so on.

The device level EMC test environment is used for devices affected by electrostatic discharge, electric field, magnetic field, the instantaneous high current and electromagnetic wave and so on.

The device and system benchmark test environment is used for devices and application systems test such as communication rate, data capacity, point to multipoint communication ability, hopping, network topology, and so on..

The air interface test environment is used for air interface test by signal identification, signal extraction and signal analysis technology with signal simulator and spectrum analyzer, and so on.

The interoperability test environment is used for application system performances test by simulating application environment and devices parameter configuration, and so on.

## 4 Summary

The performance of wireless sensing device of IOT is not only restricted by its specific air interfaces (such as frequency, modulation, protocols, and so on), but also affected by various application factors. The key of IOT application is to meet the user's needs on performances, but in an open environment, users often have to choose devices from many similar devices. Therefore, in future of IOT development, we should establish a corresponding public service evaluation center for the wireless application technology of IOT. It can provide services such as radio equipment type approval test, application scheme comparison, third party entrust test and technology consultation for the promotion, and it can offer test validation and technology assessment for the wireless devices and applications to speed up the development of IOT.

**Acknowledgments.** This work is supported by “New generation broadband wireless mobile communication network” national science and technology major project (No. 2011ZX03005-002).



## References

1. Zhang, Z.: Radio frequency identification technology theory and practice. China Science and Technology Press, Beijing (2008)
2. Zhao, J.: Radio frequency identification technology and applications. China Machine Press, Beijing (2008)
3. Ci, X., Wang, S., Wang, S.: Radio frequency identification (RFID) systems technologies and applications. Posts and Telecom Press, Beijing (2007)
4. Wang, Z.: RFID reader communication performance test platform design and realization. Beijing University of Posts and Telecommunications, Beijing (2009)

# The Selection of the Solar Battery Board Applied in New-Type Photovoltaic Generation Awnings

Xu-ping Chen\* and De-min Xiong

Department of Economic Management, Zhejiang College of Construction, Hangzhou,  
China, 311231  
jgxcxp@126.com

**Abstract.** All awnings applied in civil buildings usually have two functions; one is to resist rain and the other is to shelter from sunshine. However, the new-type awnings have the third function. The extra function is PV power generation. As the shade surface of awnings, the type of solar battery board decides whether the new-type PV powered awnings possess the above-mentioned three functions, as well as is one of the keys to whether they are worth popularizing. This text focuses on the exploration and study on the selection of solar battery board when they are used as the materials of shade surface on awnings. Based on their three basic functions, via analyses and comparisons in translucency and PV transmission rate etc. It is put forward that using thin-film solar battery board as the adumbral material is the best choice.

**Keywords:** Photovoltaic generation, awnings, solar battery board.

## 1 Introduction

The potential crises of energy and deterioration of ecological environment force all countries develop new energy and reproducible energy including solar energy [1]. New-style energy usually means nuclear energy and solar energy, wind energy and geothermal energy, hydrogen and so on. When all the new energy gets used as civilian power supply, all the security convenience, economy and operability, etc. should be considered. More over, because of its merits, solar energy will be given a considerable status again especially when nuclear energy reaches a crisis point, and the solar energy will get much wider application.

Considering the current situation of south walls of balcony and window awnings for civilian buildings, if we shift the materials into solar-powered battery board, it may come to a new photovoltaic generation awning, which will promote green energy go into the common families through making full use of solar PV generation and finally do some contribution to energy conservation and emission reduction.

---

\* Corresponding author.

## **2 Analyses on Sun-Resistant Function and Materials of Awnings for Civil Buildings**

### **2.1 Sun-Resistant Function of Awnings for Civil Buildings**

There are two functions of awnings fixed on civilian buildings. One function is to keep off the sunshine. It could not only reduce the indoor temperature and air conditioner load via keep off the roasting heat of sunshine in summer but also transfer the steaming light from great light into the diffuse light, which makes a bright room not a dazzling one. And in this way, we lengthen the service life of internal decoration and furniture. The other one is to keep off the rain. It is able to avoid internal decoration and furniture being eroded by rain even in a state of good ventilation and daylight.

This kind of awnings are mostly fixed on facades of civilian buildings, such as windows, balconies, terraces, and entrances, and this application has decades of history in architectural use.

### **2.2 Adumbral Materials of Awnings Used in Civil Buildings**

The common materials of awnings used in civil buildings are high-grade Oxford cloth, Decron Fabric, Acrylic board, PC awing board, coated steel sheet, and canvas and so on.

The common materials of building awnings have certain characters of UV-proof ability and high weather ability, security and convenience, some translucency etc. But meanwhile many problems exist such as non-ideal windproof character, great noises with big rain, and brittle material easy to chip and so on.

## **3 Feasibility Analysis on the Application of the New-Type PV Generation Awning**

These new-type PV generation awnings adopt solar powered battery boards as their adumbral materials, which add an extra function of power generation to their original two sun-proof and rainproof functions. This kind of new PV generation awnings could be set up on the dwelling buildings and office buildings as one of their important energy sources. As long as the natural sunshine is sufficient, they can take place of the partial electricity from city grids and solve the problem of electric power shortage during the current peak hours. Thus we can also cut carbon-dioxide emission as well as reduce coal and oil, etc. high-carbon energy consumption, which will finally promote the practical applications of green, eco-friendly solar energy.

### 3.1 Calculation of PV Transformation of the New-Type PV Generation Awnings

#### 3.1.1 Calculation of Available South Wall Areas for Setting Up Awnings

##### 1) Calculation of Areas of South Walls Suitable for Awnings in Common Office Buildings

Take a south-facing ordinary office as an example, its axis bay is 3.6 meters in length; the awning is 0.8 meter in width and its installed slope angle is  $15^\circ$ , so the shading areas ( $s$ ) of the office awning should be:  $s = 3.6 \times 0.8 / \cos 15^\circ = 2.98 m^2$

Consequently for a common six-floor office building, which has twelve bays per floor, its available areas ( $S$ ) for awnings could be:  $S = 6 \times 12 \times 2.98 = 358.56 m^2$

##### 2) Calculation of Available Areas of South Walls for Setting Up Awnings in Civil Multi-Storey Dwellings

A familiar dwelling built after 2000 usually has three southern rooms, sitting, sleeping and study rooms. According to the general design of the southern bays, their axis could be 4.2 m, 3.9 m and 3.3 m respectively. Therefore, its total available length for awnings is 11.4 meters. If the awning is still of  $15^\circ$  slope angle and 0.8 meter in width, the shading area ( $A$ ) of this dwelling could be

$$A = 11.4 \times 0.8 / \cos 15^\circ = 9.44 m^2$$

#### 3.1.2 Calculation of Generation Capacity of the New-Type PV Generation Awnings

##### 1) Calculation of Generation Capacity of the New-Type PV Generation Awnings Fixed on the Common Office Buildings

The shade area of the awning which is fixed on the above-mentioned common office building with a southern exposure could be up to  $2.98 m^2$ . If we choose the monocrystalline silicon battery board, and pack them according to its regular size  $156 \times 156 mm^2$ , 4.2 W, 0.55 V, we need 5 pieces in width and 22 pieces in length. Totally one awning can take 110 pieces, and its maximum power can be 462 W. Usually the areas of the above-mentioned office is no more than  $20 m^2$  so according to its luminance three 40-W fluorescent lamps are enough, and their needed electric capacity is 240. Therefore, in theory, the generating capacity of one solar awning certainly can meet the electric need of one office for illumination.

##### 2) Calculation of Generating Capacity of the New-Type PV Generation Awnings Fixed on the Residential Multi-Storey Dwellings

The above-mentioned awning adumbral areas of common civil dwelling with three rooms and two halls reach  $9.44 m^2$ . If we still choose monocrystalline silicon solar battery board, and packaging fix according to its regular size of  $156 \times 156 mm^2$ , 4.2 W and 0.55 V, we need 5 pieces in width and 72 pieces in length, totally 360 pieces. That means the maximum capacity of this awning can be 1512 W. If we calculate as average capacity of a power meter from power supply bureau, 10 (40) A, a suit of this dwelling has a 2200 W installed capacity. In other words, the output of this new style PV generation awning is able to meet its three fourths of electric use.

### 3.2 Analyses on the Influence of Solar Aawnings' Load to Building Structure

The weight of one solar awning consists of 3 parts. Part one is the weight of solar battery components; part two is the weight of the awning's frame structure; part three is the weight of brackets.

#### 3.2.1 The Weight of Solar Battery Components

##### 1) The Weight of Solar Battery Components Installed on South Single Office in Common Office Building

If we still use monocrystalline silicon solar battery board, whose regular size is  $156 \times 156 \text{ mm}^2$  in area and about 0.1 Kg per one piece of battery board in weight, we will get that the total weight of its solar battery board components is :

$$G_1 = 0.1 \times 110 = 11 \text{ kg}$$

##### 2) The Weight of Solar Battery Components Installed on Single Civil Multiple Dwellings

If we use the same solar battery board as the one used in office buildings, we will calculate the weight of solar battery board elements installed on single dwellings with three rooms and two halls as:  $G_1 = 0.1 \times 360 = 36 \text{ kg}$

#### 3.2.2 The Weight of the Awning's Frame Structure

##### 1) The Weight of the Solar Battery Awning's Frame Installed on South Single Office in Common Office Building

Technically when we pack cells into solar battery PV components, we usually use low iron tempered glass or transparent cover which is recovered by TPT material, and strengthen their circumference with aluminum alloy frame. Therefore the weight of one solar battery components' frame includes all weight of cover-used tempered glass, TPT materials and aluminum alloy frame structure materials. According to existing finished product of solar battery components, it is estimated that the components' frame structure takes up 70% of total battery components, so the weight of solar battery components' frame structure generally comes to

$$G_2 = (G_1 + G_2) \times 70 \%, \quad \therefore G_1 = 11 \text{ kg}, \quad \therefore G_2 = 26 \text{ kg}.$$

##### 2) The Weight of Solar Battery Components' Frame Structure Installed on One Set of Civil Multi-Storey Dwellings

Similarly, the weight of solar battery components' frame structure installed on one three-room and two-hall residential house is about

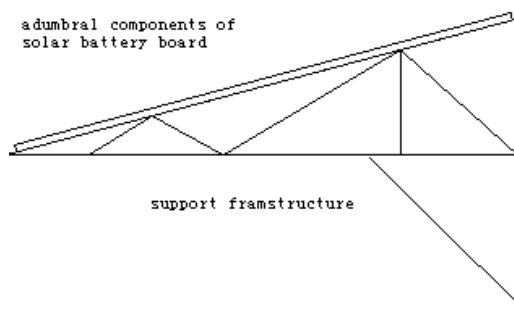
$$G_2 = (G_1 + G_2) \times 70 \%, \quad \therefore G_1 = 36 \text{ kg}, \quad \therefore G_2 = 84 \text{ kg}.$$

#### 3.2.3 The Weight of One Solar Awning's Brackets

Adumbral components of one solar awning are heavier than adumbral materials of a common awning. Compared with arilics sheets and PC awning sheets, its weight is the double of its original materials, and remains basically the same with the weight of colored steel awning board.

We may choose aluminum alloy as the material of solar awnings' brackets. Aluminum alloy features easily-worked and high durability, light in weight and low in cost, good decorative effect and rich designs and colors, etc. As the material of solar awnings' brackets, as long as the thickness of the aluminum alloy is up to or over 1.2 mm, the brackets have satisfying mechanical strength.

Considering the demand that the brackets must sustain the weight of solar battery components frame structure, we need put one set of brackets every 1.0~1.5 m. The sketch map of one set of brackets' structure is shown in Figure 1.



**Fig. 1.** Schematic diagram of support frame structure of the new type PV generation awning

It is estimated that this set of brackets made from aluminum alloy is about 1.88 kg in weight. If we put one set every 1.5 m, the maximum space, we need 4 sets for one common office, and 9 for the above-mentioned dwellings.

So we get the brackets' weight of the solar adumbral board for common office should be:  $4 \times 1.88 = 7.52$  kg

And the brackets' weight of solar adumbral board for common dwellings (three room and two halls) should be:  $9 \times 1.88 = 16.92$  kg

**3.2.4 The Influence of Solar Awnings' Load toward Building Structure**

Through calculation we get the total load of a new-style PV generation awning for one office should be :  $7.52 + 26 + 11 = 44.52$  kg. According to common value of building masonry load and wall surface load, the Table 1[2], we check it via the formula  $1 \text{ kg} = 9.8 \text{ N}$  and find the total load of the new-style PV generation awning brings no influence to the out wall load of the common office.

**Table 1.** Value of building masonry load and wall surface load

Types	Wall Materials and Practices	Load Reference
(kN/m <sup>3</sup> ) bricks and block weight	normal sintered brick with cement mortar	18
	mechanical sintered brick with cement mortar	19
	slag brick	18~19
	KP1 type sintered cork brick	14
	hollow concrete brick (390 mm×390 mm×140 mm)	9.8
	small-scale hollow concrete brick (390 mm×390 mm×190 mm)	11.8
	autoclaved fly ash aerated concrete block	5.5

Similarly, we get the total load of a new-style PV generation awning for single dwelling should be:  $16.92+84+36=136.92$  kg. Again we get the total load of the new-style PV generation awning brings no influence to the out wall load of the common dwellings via the same analysis.

## **4 Comparison Analyses of Solar Material Abilities of the New-Style PV Generation Awnings**

Via the above calculations and checking, it gets proved that the new-style PV generation awnings will not influent buildings' facades, which means those new awnings are passable to promote. However, the choice on solar materials of the awnings may directly cause differences of their three main functions, namely sun shade, rain resistance and generation ability. Therefore, on the basis of taking account of all functions, how to keep the awnings in a better shape is one of the key in their promotion and application.

### **4.1 Parameters of the Awning's Three Functions (Adumbral Ability, Rainproof Ability and Generation Ability)**

In order to make the three functions of the new awnings for a better shape and best effects in a whole, it is necessary to pick right and appropriate parameters to represent the three functions. Consequently, this provides a basis for choosing and identifying the solar materials of the new-style PV generation awnings.

#### **4.1.1 Parameter Analysis of Adumbral Ability**

To keep the hot sunshine out of the room, to lower the temperature and the work load of air conditioners, and brighten the room but weaken the glare by means of allowing diffuse light to get in are the purposes of all awnings' adumbral function.

According to their basis materials, current solar powered batteries can be classified as crystalline-silicon solar battery and non-crystalline silicon solar battery, microcrystalline silicon thin-film solar battery, polycrystalline silicon thin-film solar battery, and nanosilicon thin-film solar battery, as well as selenium cell, compound solar battery, organic semiconductor solar battery, and other things. In the aspect of their common use and convenience to draw materials, they can be grouped as crystalline-silicon solar battery and non-crystalline silicon solar battery which is also known as thin-film solar battery and crystalline-silicon solar battery includes monocrystalline silicon one and polycrystalline silicon one.

From the adumbral ability, both crystalline silicon solar battery and non-crystalline silicon solar battery are able to resist hot sunshine getting into the room and lower the temperatures, so their adumbral parameters are both excellent. Nevertheless the adumbral materials demand for some light transmittal ability if they can limit sunshine get in only in a form of diffuse light at the same time. In the consideration of this parameter, we find crystalline silicon solar battery is weak in light transmission ability, but the non-crystalline silicon solar battery does an excellent job.

Comparisons on adumbral ability parameters are shown in Table 2.

**Table 2.** Comparisons on adumbral ability parameters

Comparison Parameters Solar batteries	adumbral ability	Light transmittance	Note
Mono-crystalline silicon battery	Excellent	Weak	
Poly-crystalline silicon battery	Excellent	Weak	
Non-crystalline silicon battery	Excellent	Good	Also named thin-film solar battery

Different Materials' Adumbral Ability Parameters Comparison Table.

**4.1.2 Parameter Analysis of Rainproof Ability**

Adumbral awnings' rainproof ability mainly depends on their shape design to fulfill. As long as it features adequate extensive width and length and gets fixed in a right shear angle, the awning may have a wonderful rainproof effect. No matter what a kind of solar battery it is, monocrystalline silicon one or polycrystalline silicon one, non-crystalline silicon (thin-film) one, it reaches an excellent rainproof result in its good form design.

**4.1.3 Analyses on PV Power Generation Ability**

**Table 3.** The work abilities of solar battery components

Solar battery Comparison parameters	Mono-crystalline silicon solar battery	Poly-crystalline silicon solar battery	Non-crystalline silicon solar battery (thin-film)
Transmission rate%	14~16	12~15	6~8
Low-light performance	Board line	Board line	Good
Price	high	high	lower
Mechanical strength	Certain seismic capacity and shock resistance	Certain seismic capacity and shock resistance	Weaker seismic capacity and shock resistance
Long-term stability of the electric output	good	good	Light-induced attenuation

Comparison table of solar battery components' PV generation ability.

According to Table 3, the analyses on work abilities of solar battery components from *Household Solar PV Power Supply System* [3], written by Liu Hong, et al. and published in 2006, we may find that the crystalline silicon solar battery gains a great advantage in transmission rate, mechanical strength and long-term stability of electrical output, however in the aspects of generation ability under weak light and price, the non-crystalline silicon solar battery instead does better. Moreover the product line of non-crystalline silicon or thin-film battery, started in Jiangsu Province in 2010, has gained a considerable advantage in its lower cost and energy consumption. Compared with crystalline silicon solar battery, the non-crystalline silicon solar battery takes up only 10 % cost as the former consumes, and less silicon consumption. Therefore its transmission rate goes up to more than 10 % currently from previous 6 %.



### 4.2 Comparison between Solar Materials’ Abilities of New Style PV Generation Awnings

Gather adumbral, rainproof and generation abilities of the all above solar materials that can be used in new type awnings, we get the Table 4. Seeing Table 4, we find the three materials have their own merits and demerits. On the premise of adumbral functions fulfilled, for the adumbral use of awnings, we raise a claim in their light transmission ability. Among the PV generation parameters, mechanical strength or safety is one of the keys, while transmission rate is a question of systematic total efficiency. If we sacrifice the partial transmission rate to boost new awnings’ promotion and use, the social population effect may get greatly lifted.

**Table 4.** Comparisons of solar battery

Solar battery Comparison parameters		Mono-crystalline silicon solar battery	Poly-crystalline silicon solar battery	Non-crystalline silicon solar battery (thin-film)
Sunshade	adumbral ability	Excellent	Excellent	Excellent
	transmittance	weak	weak	weak
Rainproof ability		Excellent	Excellent	Excellent
	Transmission rate%	14~16	12~15	10
PV generation	Low-light performance	Excellent	Excellent	good
	Price	high	high	lower
	Mechanical strength	Certain seismic capacity and shock resistance	Certain seismic capacity and shock resistance	Weaker seismic capacity and shock resistance
	Long-term stability of the electric output	good	good	Light-induced attenuation

Comparison Table of adumbral ability, rainproof ability and generation ability parameters.

### 4.3 Confirmation of Solar Energy Materials Used in the New-Style PV Power Generation Awning

From the above analyses, it is concluded that the non-crystalline silicon battery, also we can say the thin-film solar battery shoulders the responsibility of the battery material used in the new-style PV power generation awning in this paper. The following are the reasons:

#### 4.3.1 Light Transmittance

Nevertheless awnings’ adumbral materials demand for some light transmittal ability so that they can limit sunshine get in only in a form of diffuse light at the same time and finally to achieve the sun-resistant job. In the consideration of this parameter, we find to use thin-film solar battery is a right choice.

#### 4.3.2 Rainproof Ability

Seeing from their rainproof abilities, the former three solar batteries all can be satisfying and remain awnings’ original rainproof function. Therefore, the thin-film solar battery may reach the requirement of rainproof function.

### 4.3.3 PV Generation Ability

We have analyzed a lot on this ability, now we make description on the PV generation ability of thin-film solar batteries.

#### Mechanical Strength

Table 4 tells that the thin-film batteries have weaker seismic capacity and shock resistance. That is because during the product process it is only allowed to use un-tempered glass. In fact for those adumbral awnings the shake and shock they need to face usually come from rains or snow, however common glass does have enough resistance and tolerance against rains or snow, such as window glass and so on.

#### Transmission Rate and Low-Light Performance

From Table 4 we find silicon solar powered batteries have the highest PV transmission rate, while the thin-film battery comparatively is in the shade. Some research shows the PV transmission rate of thin-film batteries may express 28%, but no report for daily use.

On the contrary, for another parameter, the low-light performance, the thin-film solar battery does a good in this aspect, but crystalline silicon solar powered battery does not. This indicates that the thin-film solar powered battery is most suitable for changeable and un-controlled climates. By this means, its adaptability and low light level can make up its demerit of lower PV transmission.

#### Price

Actually price is one of the constant blocks to the application and the promotion of solar powered PV generation. Actually the study results of these new awnings do cater to the mass; therefore if we want these awnings to be acceptable by common people, the price is the point. From that point, the price parameter in Table 4 shows the thin-film solar battery is the most suitable material for new-style PV generation awnings.

#### Long-Term Stability of Electrical Output

As for long-term stability of electrical output, Table 4 tells us that compared with the silicon solar powered battery or named thin-film solar battery has weakness in optical attenuation. Optical attenuation effect is a Si:H thin-film may cause inner defects after long-time strong light exposure and current pass, which makes its performance down. This is called Steabler-Wronski effect. The latest information shows that if non-silicon solar batteries' I layer in N-I-P and P-I-N structure get mixed by a little of boron, their attenuation effects get weaker greatly. In other words, the attenuation of non-silicon battery (thin-film solar battery) can get improved through choosing suitable structure and doping technique during the formation process. In this way, we are able to make thin-film solar batteries' long-term stability of electric output close to the one of crystalline silicon solar batteries, and finally meet the demands in practical use.

According to the latest research and study, non-crystalline silicon or thin-film solar battery is the best one get universally acknowledged because of its light weight, high absorption ratio, high open circuit voltage, great resistance to radiation, and good durance for high temperature, but simple preparation technique and equipment, and low consumption of energy. More over, it can be deposited on any under-layers with low temperature and short-time consumption, which means it fits for lot production and so on. [4] Consequently we pick non-crystalline silicon thin-film solar batteries as the solar battery board of new-style PV generation awnings.

## 5 Feasibility Analysis of Thin-Film Solar Generation Awning

### 5.1 Use Security

The second phrase gives the analysis and results on total load of one new type awning. We get that the new awning brings no potential safety concerns at all. That is to say, building facades can sustain the weight of new-style PV generation awnings without question. If we take thin-film solar batteries as new awning's adumbral material, the total load of the awning may be lighter than monocrystalline solar cells.

It is fully workable to fix the new type awnings with expansion bolt just as the normal way. However, the installation must be reliable. That is the key to guarantee its healthy work.

Via the analyses in the third phrase we know that the production process of the new-style PV generation awning demands higher technology. Firstly, the package size of the thin-film solar batteries' shade material should be sternly in accordance with awning specifications in encapsulation. Secondly, the framework of thin-film solar battery awnings should be firmly reliable to ensure a solid installation for their groupware. Thirdly, all metal components need fast ground handing for lighting strike.

### 5.2 Affordability

Huge investment of solar PV generation skills and new green conception of energy saving makes the fruits in this field affordable only by the rich. Usually most of the investments focus on stand-alone PV power generation system and integration system of building energy efficiency, whose expensive design fee, high-class materials and luxury decoration break common people's aspiration.

The design and study of new-style PV generation awnings are geared to the needs of the mass; therefore the price must be acceptable by common people.

If we calculate at 800 W for 6 hours as a household's average power consumption, including the 24-hour fridge's consumption, one family consumes 4.8 KW, and if we calculate at peak price 2.54 Yuan/hour, one family needs to pay 928.56 Yuan per year and 9285.6 Yuan for ten years.

The power of one thin-film solar battery is  $63.3 \text{ W/m}^2$  and its lot size price is 8 Yuan per Watt; the results from our calculation in the second phrase is a dwelling with 3 rooms and 2 halls in theory gets  $9.44 \text{ m}^2$  solar battery board and enjoys about 597 W power from the new type awning, whose lot size price is 4776 Yuan(RMB); in theory it costs less than 10 thousand Yuan even plus all control settings fee and installation fee, while it can be used more than 20 years usually.

Common people's investment for once can be reclaimed in ten years, and not blocked by shortage and restricted use of electricity. Therefore this investment can be acceptable by all common households.

### 5.3 Other Needed Settings

Through the above analyses we choose thin-film solar powered batteries to make new-style PV generation awnings' battery board, however, a complete set of the new type awning needs equipped other settings.

### 5.3.1 Controller

The new-style PV generation awnings need setting up systematic controllers. They are of the settings to prevent their storage batteries being over-charged or over-discharged.

### 5.3.2 Invertors

Illuminative equipment in civil buildings usually adopts single-phase 220 AC (alternating current) power source. The inverter plays the role of changing direct current from solar batteries' PV transmission into alternating current, and boosting the voltage to the users' power-needed equipment.

### 5.3.3 Storage Batteries

Storage batteries are one of settings in system of new-style PV generation awnings. Those solar batteries transfer solar power into direct current, and store the energy in form of chemical energy, so that they can supply stable electrical power to the load.

### 5.3.4 Gauging Table

There are two gauging tables in the system of new-style PV generation awnings. One is the purchase electric meter, which measures electricity that user gets from public grid; while the other is the sell electric meter, which registers redundant electricity that user outputs to public grid. The former one contributes the later one charges.

## 6 Peroration

Solar power is a kind of clean and green reproducible energy, which distributes its resource in an even way without limitation of countries or wealth. For our human beings, it can be taken as an inexhaustible energy, which is of great strategic significance to country energy supply and energy security. [6] It is every investigator's wish that the new-style PV generation awning may bring this green energy to common people, helping the mass get to know and understand the solar power, and finally the application of solar power gets huge promoted in this way.

**Acknowledgments.** This paper is one of the study fruit of Feasibility Study of Solar Adumbral Awnings' Promotion and Application, which is a project of Ministry of Housing and Urban-rural Construction in 2011. The project number is 2011-K1-38.

## References

1. Luo, Y.-Z., Chen, J.: Integrated Applicating of Solar Energy for Large-scale Public Buildings. *Shanghai Electric Power* 2, 142–147 (2008)
2. Load Code for the Design of Builings Structures, GB 50009-2001, pp. 30 (2001)
3. Liu, H.: Household Solar Photovoltaic Power System, p. 51. Chemical Industry Press (2008)
4. Zhang, S., Wang, X.-P., Wang, L.-J.: Research Progress on Thin film Folar Cells. *Materials Review* 5, 126 (2010)
5. Krauter, S., Wang, B.(trans.), et al.: Solar Energy Generation- Photovotaic Energy System, p. 39. China machine Press (2009)
6. Luo, Y.-Z.: Discussion on Solar Photovoltaic Technique and Long-span Building Integration. *Energy Engineering* 10(31), 145–148 (2007)

# Unbalanced Three-Phase Power Flow Calculation Based on Newton Method for Micro-Grid

Jiang Guixiu, Shu Jie, Wu Zhifeng, and Zhang Xianyong

Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, China  
{jianggx, shujie, wuzf, zhangxy}@ms.giec.ac.cn

**Abstract.** Because of the connection of distributed generation to energy complementary micro-grid, there are multi-supplying points and loop net which have serious influence on electric network tide and voltage regulation. Based on the models of different distributed generation and energy storage elements in power flow calculation, an unbalanced three-phase power flow calculation based on Newton method for micro-grid is presented, in view of three-phase unbalance and loop net of the micro-grid. The conventional steps of forming the Jacobian matrix, LU factorization and forward/back substitution in Newton method are replaced by back/forward sweeps on radial feeders with equivalent impedances, so that the calculation speed is proved. The example proved that the proposed method is efficient and feasible.

**Keywords:** Three-phase power flow, Newton method, weakly looped network.

## 1 Introduction

As the important component part of the smart grid, micro-grid traces the latest development trend of international electric power technology. The distributed energy complementary micro-grid consists of the power unit of wind energy, solar energy, and ocean energy, the diesel generating set in support, and the energy storage equipment. Integrating the distributed generation with the energy storage equipment, energy conversion device, the load, the supervision and protection unit, the micro-grid is a minor energy supply system which can either run independently, or connect to the large grid. The conventional power flow methods are not applicable to the micro-grid with multi-source system anymore, so that it's important to research on the power flow method which takes into account the influence of distributed generations.

The differences between the conventional power flow method and the method considering the distributed generations are the modeling and the treatment of distributed generation. The distributed generation was simply treated as PQ node in [1] and PV node in [2]. The familiar distributed generations come down to PQ, PV, or PQ(V) node in [3], and the three phase power flow method based on Newton method for distribution network is proposed. A three phase power flow for the weakly meshed distribution networks based on superposition principle was presented in [4]. The models of typical distributed generations such as induction generator, synchronous generator without excitation regulation, and fuel cell were built in [6], besides a

power flow for distribution networks based on sensitivity compensation was proposed.

Based on the analysis of the power output characteristic and control characteristic of different power units, combining with the distributed generation models and power flow algorithms, an unbalanced three-phase power flow calculation based on Newton method for micro-grid is presented, in view of three-phase unbalance and loop net of the micro-grid.

## 2 The Node Type of Distributed Generation

According to the energy transformation, the typical distributed generations connect to the system through synchronous generator, power electronic device, or induction generator. And the distributed generations are classified as PQ, PV, PI, P-Q(V) node based on the corresponding connection.

### 1. PQ node

The parameters of synchronous generator such as terminal voltage, reactive power, and power factor, are controlled by excitation regulator, so that when the control method of excitation regulator is constant reactive power regulation or constant power factor regulation, the distributed generation connected to system through synchronous generator can be treated as PQ node.

When the reactive output of inverter is zero, or the reactive output is controlled by constant power factor, the distributed generations such as photovoltaic generation system and energy storage equipment connected to system through inverter can be treated as PQ node.

### 2. PV node

When the control method of excitation regulator is automatic voltage regulation, the distributed generation connected to system through synchronous generator can be treated as PV node.

When inverter or converter connects to system as voltage controller, the photovoltaic generation system and energy storage equipment, or the micro turbines and parts of wind turbines, can be treated as PV node.

### 3. PI node

When the connection is power electronic device which working as current controller, the distributed generation can be treated as PI node, such as micro turbines, parts of wind turbines, photovoltaic generation system and energy storage equipment.

### 4. P-Q(V) node

The connection of wind turbines is mostly induction generator without voltage regulation, whose electromagnetic field relies on the reactive power from electric network, so that the wind turbines should be treated as P-Q (V) node.

## 3 Three-Phase Power Flow Calculation Based on Newton Method for Micro-Grid

### The Modified Newton Method

The main reason effect the calculation speed is that the Jacobian matrix should be formed and then eliminated in every iteration step, which lead to great arithmetic

labor. If the Jacobian matrix doesn't need to be explicitly formed, the speed of Newton method can be accelerated greatly.

In conventional Newton method, the power flow can finally be expressed as:

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \Delta V/V \end{bmatrix} \tag{1}$$

In modified method, because the line isn't long and the power isn't high in micro-grid, so that we can assume that there is small voltage difference between two adjacent nodes, and no shunt branches. Based on the assumptions, after being approximated, the submatrices of Jacobian matrix, H, N, J, and L all have the same properties (symmetry, sparsity pattern) as the Nodal Admittance Matrix, and they can be formed as:

$$H = L = A_{n-1} D_B A_{n-1}^T$$

$$J = -N = A_{n-1} D_G A_{n-1}^T$$

where  $D_B$  and  $D_G$  are diagonal matrices with  $p \times p$  matrix diagonal entries to be  $V_i^p V_j^m B_{ij}^{pm} \cos \theta_{ij}^{pm}$ ,  $V_i^p V_j^m G_{ij}^{pm} \cos \theta_{ij}^{pm}$ ,  $p=a, b, c$ ,  $m=a, b, c$ ,  $A_{n-1}$  is the bus to branch incidence matrix. Therefore (1) can be rewritten as:

$$\begin{aligned} \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} &= \begin{bmatrix} A_{n-1} & \\ & A_{n-1} \end{bmatrix} \begin{bmatrix} D_B & -D_G \\ D_G & D_B \end{bmatrix} \begin{bmatrix} A_{n-1}^T & \\ & A_{n-1}^T \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \Delta V/V \end{bmatrix} \\ &= UDU^T \begin{bmatrix} \Delta \theta \\ \Delta V/V \end{bmatrix} \end{aligned} \tag{2}$$

It is noted that if nodes and branches are ordered according to depth-first or breadth-first search ordering scheme,  $A_{n-1}$  is an upper triangular matrix with all diagonal entries to be 1 and all non-zero off-diagonal entries to be -1, having no need to be formed in the program.

Therefore, there is no need to form the Jacobian matrix. The calculation of each line just need its own terminal voltage, power, and line impedance, having no business with other line parameters.

$$E = \Delta \theta + j \Delta V/V$$

Define  $S = \Delta P + j \Delta Q$ , then (2) can be written as:

$$W = D_B + j D_G$$

$$A_{n-1} W A_{n-1}^T E = S \tag{3}$$

$$A_{n-1} S_L = S \tag{4}$$

$$W A_{n-1}^T E = S_L \tag{5}$$

Where (4) is the back sweep for  $S_L$ , and (5) is the forward sweep for E. An algorithm based on (4) and (5) has been derived.

**The Solution of Loops**

Assuming node  $k$  is the selected breakpoint in a loop, and it is broken into nodes  $i$  and  $j$ , then there are boundary conditions as following:

$$\begin{aligned} \Delta\theta_i - \Delta\theta_j &= 0, & \Delta V_i - \Delta V_j &= 0 \\ \Delta P_i + \Delta P_j &= \Delta P_k, & \Delta Q_i + \Delta Q_j &= \Delta Q_k \end{aligned}$$

After performing one row operation on  $U$  and one column operation on  $U^T$ ,  $UDU^T x = b$  can be transformed into the form as  $Cx_b = b_x$ , and  $x$  in  $x_b$  and  $b_x$  can be obtained, where subscript  $b$  in  $x_b$  means some of the entries in  $x_b$  are known and similarly, subscription  $x$  in  $b_x$  means some of the entries in  $b_x$  are unknown. Note the number of unknown variables in  $b_x$  is the same as the number of known variables in  $x_b$ . Also, these known and unknown pairs are on the same rows of  $x_b$  and  $b_x$ .

**The Treatment of Distributed Generation in Power Flow**

1. PQ node

The distributed generation of PQ node type can just be treated as negative load. The constant reference value of active power  $P_i^p$  and reactive power  $Q_i^p$  in this step are needed, and  $\Delta S^{(k)}$  of  $k$ -th iteration can get:

$$\begin{aligned} \Delta P_i^p &= P_i^p - \left| U_i^p \right| \sum_{j=1}^n \sum_m \left| U_j^m \right| \left( G_{ij}^{pm} \cos \theta_{ij}^{pm} + B_{ij}^{pm} \sin \theta_{ij}^{pm} \right) \\ \Delta Q_i^p &= Q_i^p - \left| U_i^p \right| \sum_{j=1}^n \sum_m \left| U_j^m \right| \left( G_{ij}^{pm} \sin \theta_{ij}^{pm} - B_{ij}^{pm} \cos \theta_{ij}^{pm} \right) \\ i &= 1, 2, \dots, n-1; p=a, b, c; m=a, b, c \end{aligned}$$

2. PI node

The PI node injection reactive power should be calculated before every iteration, and the PI node can be treated as PQ node with active power  $P$  and reactive power  $Q_{k+1}$  in the  $(k+1)$ th iteration. The reactive power  $Q_{k+1}$  can be solved by the following formula:

$$Q_{k+1} = \sqrt{|I|^2 (e_k^2 + f_k^2) - P^2} \tag{6}$$

where  $I$  is the constant current amplitude,  $e_k$  and  $f_k$  are the real and imaginary parts of voltage ( $e_k + jf_k = U_k$ ).

3. P-Q(V) node

The P-Q(V) node voltage should be modified after every iteration, then the absorption reactive power and power factor can be calculated. According to the demand to power factor of node, the number of input shunt capacitor bank can be got,



and the reactive power compensation in fact is figured out. The difference of the absorption reactive power and the reactive power compensation is the absorption reactive power of the node, hence the P-Q(V) node can be treated as PQ node.

Define the P-Q(V) node output power  $P_e$  is the output power of induction generator, U is the modified node voltage, the absorption reactive power of the node Q can be calculated as following:

$$s = \frac{r(U^2 - \sqrt{U^4 - 4x_\sigma^2 P_e^2})}{2P_e x_\sigma^2}, \quad Q' = \frac{r^2 + x_\sigma(x_m + x_\sigma)s^2}{rx_m s} P_e$$

$$Q_c = P_e \left( \sqrt{\frac{1}{\cos^2 \phi_1} - 1} - \sqrt{\frac{1}{\cos^2 \phi_2} - 1} \right), \quad [n] = Q_c / Q_{N-Unit}$$

$$Q'' = nQ_{N-Unit} U^2 / U_N^2, \quad Q = Q'' - Q'$$

where s is the slip of induction generator, r is rotor resistance,  $x_\sigma$  is the sum of stator reactance  $x_1$  and rotor reactance  $x_2$ ,  $x_m$  is the excitation reactance,  $Q'$  is the absorption reactive power of induction generator,  $\cos \phi_1$  and  $\cos \phi_2$  are the power factor of induction generator and node with shunt capacitor banks,  $Q_c$  is the needed reactive power compensation of shunt capacitor banks, n is the number of shunt capacitor banks,  $Q_{N-Unit}$  is the reactive power compensation of each shunt capacitor bank,  $Q''$  is the reactive power compensation of shunt capacitor banks in fact.

4. PV node

The submatrices of Jacobian matrix, J and L have not the row corresponding to PV node, and the Jacobian matrix can't be transformed to the symmetry pattern as (2), so the proposed method can't deal with the PV node. Therefore the injection reactive power demand that can maintain the node voltage should be worked out as the reference value of PV node reactive power. And the PV node can be treated as the PQ node. The injection reactive power  $Q_j^p$  can be got as following:

$$Z_{PVj}^p \dot{I}_{compj}^p = \Delta \dot{V}, \quad I_{compj}^p = \frac{\Delta V_{PVj}^p}{|Z_{PVj}^p|}$$

$$Z_{PV} = \begin{bmatrix} |Z_{PV1}| & \cdots & |Z_{PV1n}| \\ \vdots & \ddots & \vdots \\ |Z_{PV1n}| & \cdots & |Z_{PVn}| \end{bmatrix}, \quad \dot{I}_{compj}^p = I_{compj}^p \times e^{j90+\delta_j^p}$$

$$Q_j^p = \text{Im}(\dot{V}_j^p \times \dot{I}_{compj}^p)$$

$$= \begin{cases} Q_j^p & \text{if } \forall_{p=a,b,c} Q_j^p \leq Q_j^{p \max} \\ Q_j^{p \max} & \text{if } \exists_{p=a,b,c} Q_j^p \geq Q_j^{p \max} \end{cases}$$

where  $\Delta V_{pvj}^p$  ( $p = a, b, c$ ) is the p-th phase difference of the J-th node,  $I_{compj}^p$  is the each phase compensate reactive current amplitude of every PV node,  $Z_{pvj}^p$  is the impedance of branch connecting j-th PV node to the source node,  $Z_{pv}$  is the PV sensitivity matrix,  $|Z_{pvi}|$  is the sum impedance amplitude of branches connecting i-th bus to the source node,  $|Z_{pvij}|$  is the repeated part of branches connecting i-th and j-th PV node to the source node,  $\hat{i}_{compj}^p$  is the compensate reactive current phasor,  $\delta_j^p$  is the p-th Voltage Phase-Angle of the PV node.

### 4 Test Results

A 33 nodes system with obvious loops and unbalanced three phase representation is taken as test system, whose detailed topology and original data are presented in [10]. The rated high side voltage is 12.5kV, the low side voltage is 0.38kV, the transformer ratio is 12.5/0.38, and the impedance conversion of low side voltage is  $0.01 + j0.04\Omega$ . Two loops of the system are taken. The three phase photovoltaic generation system treated as PI node is connected at node 34, the distributed generation treated as PQ node(P=300KW、 Q=100kvar) is connected at node 36, the induction generator(P=300KW、 s=0.03) treated as P-Q(V) node is connected at node 38, the distributed generation treated as PV node(P=500KW、 U=12.5kV) is connected at node 39.

The proposed method is used for power flow calculation, compared with the conventional Newton method. The convergence precision is  $10^{-4}$ . Tab.1 shows a part of results of node voltage. Tab.2 shows the time used and iteration number.

**Table 1.** Results of node voltage

<b>node</b>	<b>1</b>	<b>6</b>	<b>12</b>	<b>18</b>	<b>22</b>	<b>25</b>
Newton method	12.5	12.1377	11.7606	11.1352	12.4608	12.3385
Proposed method	12.5	12.1377	11.7606	11.1352	12.4608	12.3385
<b>node</b>	<b>36</b>	<b>38</b>	<b>39</b>	<b>40</b>	<b>41</b>	
Newton method	12.3385	0.1722	11.8965	12.5	0.1722	
Proposed method	12.3385	0.1723	11.8965	12.5	0.1723	

**Table 2.** Computational performance comparison of two methods

	<b>Newton method</b>	<b>Proposed method</b>
iteration number	12.5	12.1377
calculating time	12.5	12.1377

From Tab.1 and Tab.2, it is seen that the proposed method can solve power flow for unbalanced three-phase micro-grid with distributed generations and loops. The traditional Newton method has better convergence. However, the proposed method has higher calculation speed, which can meet the demand of speediness.

## 5 Conclusions

An unbalanced three-phase power flow calculation based on Newton method for micro-grid is proposed, in which the Jacobian matrix can be formed as  $UDU^T$  under certain assumptions, where  $U$  is a constant upper triangular matrix depending solely on system topology and  $D$  is a block diagonal matrix. So that the Jacobian does not need to be explicitly formed, and a back/forward iteration algorithm based on the linearized power flow equations can be used instead of the conventional LU factorization. Thus, the possible ill-conditions associated with the Jacobian matrix and its LU factors are completely avoided. Besides the treatments of distributed generation in power flow are discussed, that is to say how to change the various type node into the PQ node or PV node.

The test results shown that the proposed method can deal with the system with loops and unbalanced three phase representation, has good convergence property and higher calculation speed.

**Acknowledgments.** The authors are grateful for the financial support from Knowledge Innovation Program of The Chinese Academy of Sciences (No. KGCX2-YW-399+8); The Science and Technology Project of Guangdong Province and Ministry of Education of China (No. 2010A090200065); Guangzhou science and technology plan projects (NO. 2010U1-D00231).

## References

1. Abur, A., Singh, H., Liu, H., et al.: Three phase power flow for distribution systems with dispersed generation. In: 14th PSCC, Sevilla, vol. 11(3) (2002)
2. Cheng, C.S., Shirmohammadi, D.: A three-phase power flow method for real-time distribution system analysis. *IEEE Trans. on Power Systems* 10(2), 671–679 (1995)
3. Wang, S., Huang, L., Wang, C.: Unbalanced three-phase power flow calculation for distributed power generation system. *Electrical Power Automation Equipment* 27(8), 11–15 (2007) (in Chinese)
4. Che, R., Li, R.: A new three-phase power flow method for weakly meshed distribution systems. In: *Proceedings of the CSEE*, vol. 23(1), pp. 74–79 (2003) (in Chinese)
5. Mishimav, Y., Sugawara, K., Satoh, T., et al.: Three-phase power flow for friends network. In: *ISCAS 2005*, vol. 5, pp. 4733–4736 (2005)
6. Chen, H., Chen, J., Duan, X.: Study on power flow calculation of distribution system with DGs. *Automation of Electric Power Systems* 30(1), 35–40 (2006) (in Chinese)
7. Ding, M., Guo, X.: Three-phase power flow for the weakly meshed distribution network with the distributed generation. In: *Proceedings of the CSEE*, May 2009, vol. 29(13), pp. 35–40 (2009) (in Chinese)
8. Zhang, F., Cheng, C.S.: A Modified Newton Method for Radial Distribution System Power Flow Analysis. *IEEE Transactions on Power Systems* 12(1), 389–397 (1997)

9. Wang, C., Zheng, H., Xie, Y., et al.: Probabilistic power flow containing distributed generation in distribution system. *Automation of Electric Power Systems* 29(24), 39–44 (2005) (in Chinese)
10. Che, R., Li, R.: A new three-phase power flow method for weakly meshed distribution systems. In: *Proceedings of the CSEE*, vol. 23(1), pp. 74–79 (2003) (in Chinese)
11. Wang, S., Jiang, X., Wang, C.: Power flow analysis of distribution network containing wind power generators. *Power System Technology* 30(21), 42–45 (2006) (in Chinese)
12. Chiradeja, P., Ramakumar, R.: An approach to quantify the technical benefits of distributed generation. *IEEE Trans. on Energy Conversion* 19(4), 764–773 (2004)
13. Wang, Z., Zhu, S., Zhou, S.: Impacts of distributed generation on distribution system voltage profile. *Automation of Electrical Power Systems* 28(16), 56–60 (2004) (in Chinese)
14. Feijóo, A.E., Cidrás, J.: Modeling of wind farms in the load flow analysis. *IEEE Transactions on Power Systems* 15(1), 110–115 (2000)
15. Tong, S., Miu, K.N.: A network-based distributed slack bus model for DGs in unbalanced power flow studies. *IEEE Transaction on Power System* 20(2), 835–842 (2005)
16. Naka, S., Genji, T., Fukuyama, Y.: Practical equipment models for fast distribution power flow considering interconnection of distributed generators. In: *IEEE Power Engineering Society Summer Meeting, Panel on Challenges on Distribution System Analysis*, Vancouver, Canada, pp. 1–6 (2001)
17. Chen, T.-H., Chen, M.-S., Inoue, T., et al.: Three-phase co-generator and transformer models for distribution system analysis. *IEEE Transaction on Power Delivery* 6(4), 1671–1681 (1991)

# Research of the Design of a Scene Simulation General Framework

Yao Linhai, Kang Fengju, and Han Hong

Marine College of Northwestern Polytechnical University, Xi'an, China; Underwater Information and Control national Lab, Xi'an, China,

Phone:+8613572098326

cxfleat@163.com, kangfengju@nwpu.edu.cn, hanhong\_1984@163.com

**Abstract.** Aiming at the issue that occurs many repeated work when developing scene simulation system, proposed a method of the scene simulation general framework design, and designed the framework including scene edit, scenario parse, initial configure, model drive and viewpoint control; aiming at the demand of the componentization of the entities in the scene simulation, proposed a method of component-based 3D entities modeling, and realized the dynamic loading and special driving of each component by the technology called configuration table and memory mapping. Through the development and application of a typical surface warfare scene simulation system, proved the practicability of the general framework.

**Keywords:** Scene simulation, general framework, component- based modeling, model driving.

## 1 Introduction

In the past, developing a scene simulation system always need to pass through a series of process such as system construct, function design and program compile, there're many repetitive work in this process. One scene simulation general framework will avoid the repetitive work during the developing of the scene simulation system, and the developer will fix attention on realize the special demand of the scene simulation system and develop the efficiency. So propose a method of scene general framework designing. This method includes scene editing, scenario parsing, original option, model driving and viewpoint controlling.

Moreover, the fore passed modeling method always aimed at single entity based on macrocosm, and led to components of the models can not directly receive the exterior data and be driven as an unattached entity, such as carrier radar, artillery and missile. The movement of the components was just simple modal expression. But along with the development of the armament simulation, each armament simulation node of the distributed alternant campaign antagonize simulation system is provided with the mathematic model which is comparatively separated from the platform level entity. The design of scene general framework can fully realize the campaign scenarios of various distributed alternant campaign antagonize simulation system by standardizing the scene developing flow and the general interface designing.

In this method, component-based 3D modeling technology means to build entity model through the component-based mode, express and set up the various armaments that carried by platform level entity as components form. Each component model can accomplish dynamic loading and data driving by receiving data sent from the relevant armament simulation node.

## 2 Total Architecture of the General Framework

Scene general framework should can neatly and fleetly organize a campaign scene simulation system. As also, the framework should have the characteristic include well reusability, expansibility, integration, generality and friendly user interface.

The process of developing a scene simulation system mainly is: create scene; parse scenario file; initially configure simulation entity; receive data and drive model.

According to the function and work flow, split the general framework into six modules: component-based models base, scene editing, scenario parsing, initial configuration, model driving and viewpoint controlling. The designing of the framework is expressed in figure 1.

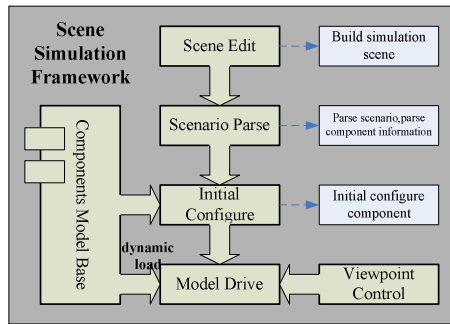


Fig. 1. Scene simulation general framework

## 3 General Framework Function Modules

### Component-Based Models Base

#### (1) Component-based modeling method

Component-based 3D modeling technology means to build entity model through the component-based mode, express and set up the various armaments that carried by platform level entity as components form. Each component model can accomplish dynamic loading and data driving by receiving data sent from the relevant armament simulation node.

Component-based 3D modeling method is described as follows:

① Aim at the components that are fixed with the platform, build detailed components node in the platform models through using the technologies such as DOF, SWITCH and ANIMATION.

② Aim at the components that are not fixed with the platform, build the un-dependent models.

During the simulation, if only need observe the movement and interaction of the entity on platform level, then not have to pay attention to the components, select the models of platform-level granularity, in this way, it can save on the system resource. If need drive the equipment components solely, well then select the models of equipment or component level granularity.

Aim at the components that move separately with the platforms, use the method called sub-model replacing that means to provide the models of two or more granularity level for user to a specific problem in the right time. For example, when the missile is launching from the ship, the change of the relative position between the component and the platform is simple linear movement, so this process can be realized by the technology DOF, after the launching process end, the missile will move freely relative to the ship, it's component 3D model will be loaded in the scene, and be driven through receiving the data including position coordinate, attitude angle and state flag etc.

(2) Foundation of the component-based models base

① Foundation of component model

For satisfying the demand of uniform management, adopt several standard rules:

- a) Uniform naming rule of component ID;
- b) Uniform coordinate;
- c) Uniform scaling;
- d) Uniform grading standard of LOD.

The step of building the component model is as follows:

- a) Build whole model of the component, and name its ID;
- b) Aim at the plane direction, create DOF1 node;
- c) Aim at the pitching direction, create DOF2 node.

② Component configuration table

The component configuration table that is expressed by XML format includes carried platform, type and replacing sub-model etc attributes, as figure 2. During the simulation, components will be driven through reading the configuration table.

```
<?xml version="1.0" encoding="GB2312" ?>
- <Component>
  <Component1 ID="component1"
    Platform="Xtype" Type="artillery"
    Model="Ytype" DOF1="RotateDof1"
    DOF2="PitchDof1" ReplaceModel=""
    PositionX="20" PositionY="20"
    PositionZ="20" />
  <Component2 ID="Component2"
    Platform="Xtype" Type="missile"
    Model="Ytype" DOF1="" DOF2=""
    ReplaceModel="Compo.fit"
    PositionX="20" PositionY="20"
    PositionZ="20" />
  <Component3 ID="Component3"
    Platform="Xtype" Type="radar"
    Model="Ytype" DOF1="RotateDof3"
    DOF2="PitchDof3" ReplaceModel=""
    PositionX="20" PositionY="20"
    PositionZ="20" />
</Component>
```

Fig. 2. Component configuration table

Aim at the components that move together with the platform, it's no need to replace sub-models, DOF1 and DOF2 respectively express the rotation and pitching DOF node. Aim at the components that move discretely with the platform, it's need to configure the FLT file's name of replacing sub-models.

### Scene Editing

The scene edit can quickly set up simulation scene through user interface form.

The scene is made up of man-made scene and natural scene. The natural scene includes the sun, the moon, the stars, clouds, rain, wind, and snow etc various natural factors. The man-made scene mainly describes terrain of the real world, characters of the earth's surface, characters of civilization, and man-made models.

The various elements of the scene are organized by a tree structure which can be edited. According to a certain format, generate the interface configuration file of application system through organizing environment, model object, observer, channel, window and pipeline etc, and then, generate the XML table which will be provided to the application system through XML parsing.

### Scenario Parsing

Convenient for initial configuration, the scenario parsing load the XML formatting simulation scenario file, and read the data information of various entities, then generate corresponding data structure. Store the information of entities including ID, type, position, attitude and state etc into a configuration file which has settled formatting. For the component searching and driving, store the affiliation, relative position, movement state etc information into the component configuration table.

Because the formats of scenario files is various, the parsing method aiming at the various scenarios must be redesigned if use groovy parsing technology, so, select the technology ADO.NET for parsing scenarios. Found a scenario template that has standard format, first transform the scenario file into the standard scenario template before parsing the scenario, then begin to parse. In this way, the problem which is generated by the reason of the various formats of the scenario files will be avoided.

The steps of the scenario parsing based ADO.NET are described as follow:

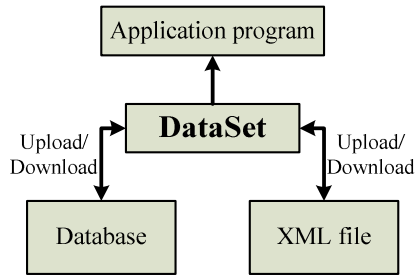
① Write the XML formatting scenario file into the DataSet objects by the data accessing technology of ADO.NET;

② Analyze the datasheet and relation of the DataSet objects namely the entities and their organization of the scenario, store the data of the DataSet into the database of scene simulation system through operating the database of ADO.NET;

③ According to the configuration information of the weapon entities that be stored in simulation database, select the relevant 3D models of the 3D models base, and associate the special effect parameters that correlate the entities.

DataSet is a data former which be in memory. It's not database, but it also include datasheet and data relation, and all of these objects are memory objects. The data included by DataSet may be from database, also may be from XML file, as the figure 3.





**Fig. 3.** Relationship of DataSet, XML file and Database

The key of accessing the XML file by technology ADO.NET rests with loading the XML file into the DataSet objects or writing the DataSet into the XML file. The functions of accessing the XML file packaged in the DataSet class are as follows table1.

**Table 1.** Operation functions of XML

Functions	Signification
ReadXml(String)	Load the data and structure of the specified XML file into the DataSet.
ReadXmlSchema(String)	Load the structure information of the specified XML file into the DataSet.
WriteXml(String)	Write the content of the DataSet into the specified XML file.
WriteXmlSchema(String)	Write the DataSet structure of the XML structure formatting into the XML file.

### Initial Configuration

Initial configuration means to initially configure the entities of platform-level or component-level according to the result of scenario parsing. Load the models from the component-based 3D models base and assemble the component-based models according to the component configuration information, then, load the models into the simulation scene and configure them to the initial position.

### Model Driving

#### (1) Searching component

The platforms and components of scene simulation are all dynamically generated according to the scenario file, and aiming to the different scenarios, need to define a general data matching scheme. Because an entity's ID is settled and only in the process of once simulation, so, use the entity's ID as matching gist, and match data through the memory mapping. During the simulation, the scene node receives data and stores the data into shared memory field, then matches and distributes the data through searching the component configuration table according to the ID of the platform and component.

(2) Component-based model driving

According to type the entities in the simulation scene are partitioned to platform type and component type. Design a entity class including the attributes and operations that all the entities have in common, such as position, attitude etc. Base on the entity class as base class, create platform class (such as ship class, plane class) and component class and regard the classes as child class of the entity class. They inherit all the attributes and operations of entity class, at the same time, each of them have special attributes and operations of themselves. The components objects of the platforms that have different ID transfer the variables and functions of these classes according to the type, and drive component-based models by the received simulation data. The flow of driving component-based models is as follows figure4.

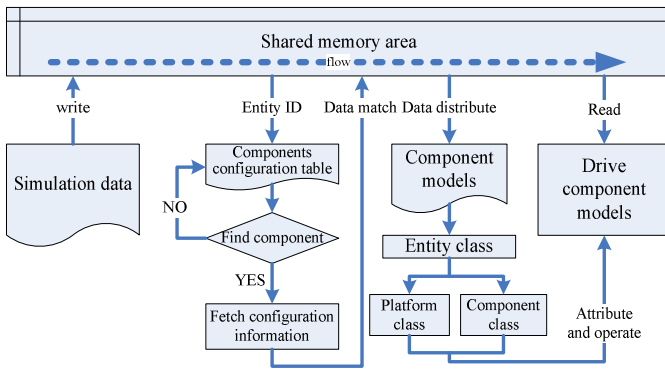


Fig. 4. Flow of driving component-based models

**Viewpoint Intelligent Control**

Import the thinking of Agent into the viewpoint control. Through making use of the basic characters of Agent, make the viewpoint of the scene simulation can automatically switch observing objects according to the hotspot events in the scene, and can automatically adjust the observing position and angle.

Create three kinds of Agent: hotspot events search Agent (Agent-HES), select observing target Agent (Agent-SOT) and select observing mode Agent (Agent-SOM). In the simulation, the viewpoint intelligent control will be realized by the cooperation of the hotspot events search Agent (Agent-HES), select observing target Agent (Agent-SOT) and select observing mode Agent (Agent-SOM), as follows figure5.

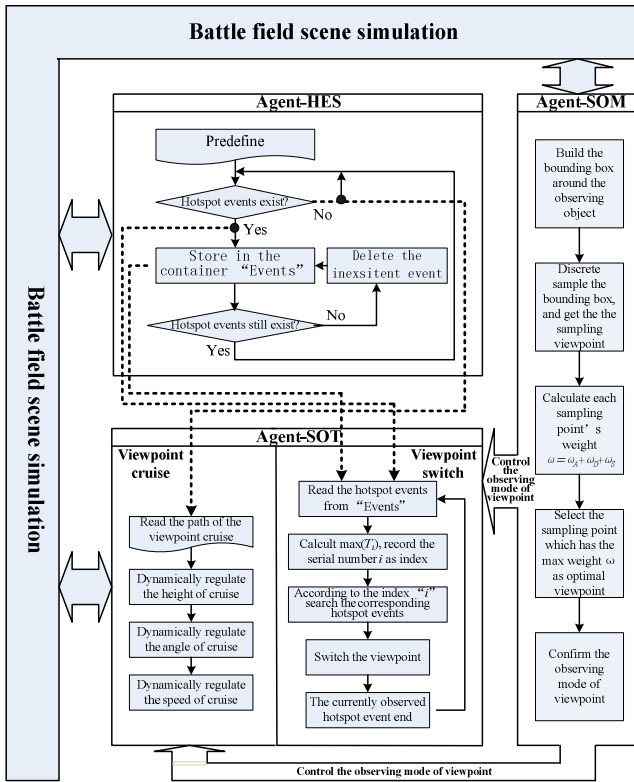


Fig. 5. Intelligent control of viewpoint based on multi-Agent

Agent-HES answers for search the hotspot events, and sends message to Agent-SOT that if there are hotspot events exist in the scene, in the same time, stores the position coordinates etc correlative information of the happened hotspot events of the scene into hotspot events container “Events”.

Agent-SOT answers for select observing objects during the simulation, includes two modes, viewpoint cruise and viewpoint switch. When the Agent-HES sends message to Agent-SOT that there are not any hotspot events in the scene, the Agent-SOT select the viewpoint cruise mode; on the contrary, when the Agent-HES sends message to Agent-SOT that there are hotspot events in the scene, the Agent-SOT select the viewpoint switch mode. When Agent-SOT is in the viewpoint cruise mode, the selection of the observing objects will be decided by the prearranged cruise path of viewpoint, and dynamically regulates the height, angle and speed of viewpoint’s cruise. When Agent-SOT is in the viewpoint switch mode, hotspot events will be read from the hotspot events container “Events”, and the observing objects of viewpoint will be selected by the calculating the PRI exponential of the hotspot events.

Agent-SOM answers for select viewpoint’s observing mode including the viewpoint’s position, the line of sight etc. Discretely sample the bounding box of observing object, and calculate the weights of these sampling points, then, the point which has the max weight will be the optimal observing position, in this way, the

observing mode of viewpoint will be controlled. During the simulation, whether Agent-SOT is in the viewpoint cruise mode or the viewpoint switch mode, Agent-SOM always use the same method to control the observing mode of observing objects.

### 4 Application Instance

Use a typical surface warfare simulation as application instance, in this simulation system, the various weapons simulation, scene simulation, environment simulation and situation evaluation etc nodes of both the red and blue sides interact and base on HLA technology. The character of these weapons presents multilayer and multi-granularity. There are ship and plane etc platforms as well as artillery and missile etc components. So split the granularity into platform level, equipment level and component level, as figure6.

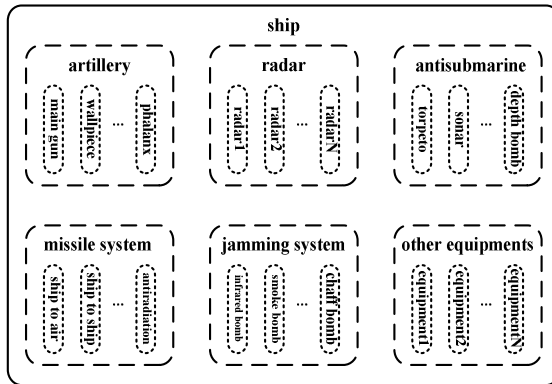


Fig. 6. Multi-granularity level of the 3D modeling of one ship

The platform level granularity is the whole model of the ship; the equipment level granularity models include artillery system, radar system, missile system and jamming system, it has not to describe the details of these equipments; the component level granularity models need to embody the idiographic equipment component carried on platform.

Use one type of artillery to show the building of component model:

- ① Build the holistic model, name the ID as “Canon”;
- ② Because the artillery can rotate freely, so generate a DOF node named “CanBotDof”, put the artillery “Canon” into this DOF node;
- ③ Because the gun tube can pitch, so also generate a DOF node named “CanTubeDof” as child node of the “CanBotDof”.

In the simulation, the rotation and pitching of this artillery can be controlled by “CanTubeDof” and “CanBotDof”.

The component configuration file is as figure7.

```

<?xml version="1.0" encoding="GB2312" ?>
- <Component>
  <Component1 ID="Artillery"
    Platform="XtypeShip" Type="artillery"
    Model="Ytype" DOF1="CanBotDof"
    DOF2="CanBarDof" ReplaceModel=""
    PositionX="20" PositionY="20"
    PositionZ="20" />
  <Component2 ID="Missile"
    Platform="XtypeShip" Type="missile"
    Model="Ytype" DOF1="" DOF2=""
    ReplaceModel="Missile.flt"
    PositionX="30" PositionY="30"
    PositionZ="20" />
  <Component3 ID="Radar"
    Platform="XtypeShip" Type="Radar"
    Model="Ytype" DOF1="RadTurnDof"
    DOF2="RadPitchDof" ReplaceModel=""
    PositionX="10" PositionY="25"
    PositionZ="20" />
</Component>

```

Fig. 7. Component configuration table of this instance

Realize the surface warfare scene simulation system according to the scene general framework, the process as follows:

- ① Build the component-based models of entities and build the component-based models base;
- ② Build the holistic framework of the surface warfare scene simulation system according to the scene general framework;
- ③ Compile the program of each module of the surface warfare scene simulation system;
- ④ Design and develop the specially required functions of the surface warfare scene simulation system and embed them into the framework;
- ⑤ Join the surface warfare scene simulation system into the distributed interactive system, debug and maintain.

Some effects of the system realizing are as figure8, figure9 and figure10.



Fig. 8. Flying F16



Fig. 9. Firing artillery



Fig. 10. Chaff cloud

## 5 Conclusion

Through the developing of the typical surface warfare scene simulation system, prove the scene general framework can effectively avoid repetitive work, make the developers can pay more attention to the designing and developing of the specifically requirements of the idiographic system, thereby advance the efficiency of developing. And, the scene simulation system based on the component-based models can satisfy

the requirement that independently drive the components in the distributed interactive simulation, make the scene simulation be more real and veracious.

Besides, the developed component-based 3D models have favorable reusability. Aiming at the component-based modeling requirement, for creating a new model, it only needs to define the carried relation, relative position and DOF of sub-model.

**Acknowledgments. Foundation item:** this work is supported by the Project of Underwater Information Processing and Control National Lab Foundation(9140C2305041001).

**Contact author:** Yao Linhai(1980), male, Ph.D. candidate, research interests includes system simulation and virtual reality.

**Mail address:** Northwestern Polytechnical University 32#, Xi'an 710072, China.

## References

1. Wang, L., Meng, X., Zhong, B.: Design and Realization of Flight Scene Simulation System. *Journal of System Simulation* (1) (2008)
2. Pan, B., Tang, S.: Design and Implement of Assembling Visual Flight Simulation System. *Journal of System Simulation* (2) (2007)
3. Wang, J., Kang, F., Liu, M.: The Technology for Rebuilding 3D Models by Components. *Computer Simulation* (10) (2009)
4. Chu, Y., Tang, S.: Research of Extensible Scene Visualization System Based on Vega Prime. *Computer Engineering and Design* 30(17) (2009)
5. Zhai, H., Cao, W., Cheng, J.: Realization of Viewpoint Control Algorithm Using Trigonometric Function. *Computer Engineering and Design* 30(7) (2009)

# Research on the Prediction of Web Application System Aging Trend Oriented to User Access Intention

Jun Guo, Hao Huang, Bo Wang, and Yunsheng Wang

College of Information Science and Engineering, Northeastern University Shenyang, China  
guojun@ise.neu.edu.cn

**Abstract.** The phenomenon of software ageing is very common with the increasing complexity of computer application system. In order to execute the software regeneration better and confirm the optimal time for software regeneration, this thesis puts forward a user access intention- oriented aging prediction method of web system. This method regards the user load as the research object, and indirectly predicts software aging by predicting the user access intention. The experimental results show that this strategy has good usability and validity, and it can predict the user follow-up visit sequence exactly. This method also laid a foundation for further execution of software regeneration.

**Keywords:** Software aging, User access intention, User pattern mining, Pattern matching.

## 1 Introduction

With the growing application range of the computer system, the functional requirement for the computer system is also increasing. Systems, long-running steadily working, become more important for the enterprises. Related research shows that the problems such as failure appear in operation process of the computer system are mainly caused by software aging. In the long-term continuous operation process, software appears the system performance down even hang down, the phenomenon is called software aging [1]. The reasons why software aging happens are software exist some problems in the long-term continuous operation and in the process of software development, the internal hidden bugs are difficult to find and modify, absolutely avoid software aging is impossible. [2]

The software anti-aging strategy proposed by Y. Huang in 1995 is the generally technology that is used to process software aging [3, 4]. Software regeneration refers to by stopping the program periodically to make the system return to normal state. The key issue is the choice of the optimal time of regeneration, the prediction of software aging trend is a crucial step. The prediction of software aging trend generally can be grouped into two basic methods: one method is time-based software aging trend prediction, the key point is to build mathematical model and predict the aging trend according to the past running state of the software system. In the literature [5], Y. Huang proposed a simple Markov model based on continuous-time. The other

method is measurement-based software aging trend prediction. [6] The key point is to monitor some performance parameters during the actual process of software running, and then predicting software aging degree according to the current period of time monitored parameters. In the literature [7], S.Garg proposed measurement-based software aging trend prediction for the first time, this method uses SNMP-based distributed monitor tool to get the key parameter from the system, and employs statistical methods to calculate, and then finally uses different strategies to decide when to execute software regeneration measures. The time-based method studies problems of software aging from a macro point of view; this method only cares the overall of software aging. Measurement-based method studies problems of software aging from a microscopic point of view, by real-time monitoring the system parameters that may cause software aging to predict the trend of software aging.

For load frequently changing web applications, measurement-based method has some deficiencies, it only through the server system parameters to monitor software aging, and it does not take the changes of system load into consideration. This thesis puts forward a user access intention-oriented prediction method of web system aging trend based on the measurement-based software aging trend method, this method regards the user loads as the research object, the main idea is: calculating the amount of damage of the web server aging by calling the page, establishing a user access patterns library, predicting the user follow-up access sequence through matching the current user access sequence with the pattern library, then predicting the system software aging. Finally, this paper proves the accuracy, effectiveness and feasibility of the proposed method by the user access intention-oriented prediction experiment and web system the aging trend prediction experiment.

## 2 The Prediction Method of Web System Aging Trend Oriented to User Access Intention

The basic ideas of the prediction method of web system aging trend oriented to user access intention are to predict the future intention of the user access according to the user's requested pages to the server in period of time in the web environment, then accumulating the follow-up visits of all users to get the visits of each page next period of time. The whole prediction process can be divided into two phases. In preparation phases, first, calculate the amount of the overall aging damage in a period of time, and calculate the amount of aging damage of each page called each time in application server. Then mine the frequently visit sequence, the access time of the adjacent page in frequent sequence patterns from user historical access log [8]. Finally, put the amount of the page damage, user frequently access sequence and access time together to get the user access patterns. In implementation phases that is to achieve the purpose of predicting user access intention by matching real-time user access sequence and the use access pattern library, and the pattern matching is to predict an individual user access sequence in the next period of time.

Figure 1 shows the overall predicting process. From the overall map of software aging process can be seen that, to get the user's access pattern library, we need the calculated outputs of three models, respectively are the page damage calculation model, the sequential pattern mining model and the access time statistical model.



The main idea of calculating the amount of page damage is to calculate the amount of the overall system aging damage in a period of time according to the recorded system server status and the software aging calculation model, then use the results and page aging damage calculation model to calculate the amount of aging damage of each page called each time in application server. Sequential pattern mining model is to mine the frequently visit sequence from user historical access logs, so as to determine the amount of user access damage and prepare for the later work. The output of the access time statistical model is the access time of the adjacent pages in frequent pattern.

### 3 The Establishment of User Access Patterns Library

#### The Preprocessing of Web Log

The data format in the log file which generated when users visit web site is not suitable for further usage, so the data of web log should be preprocessed. Its main

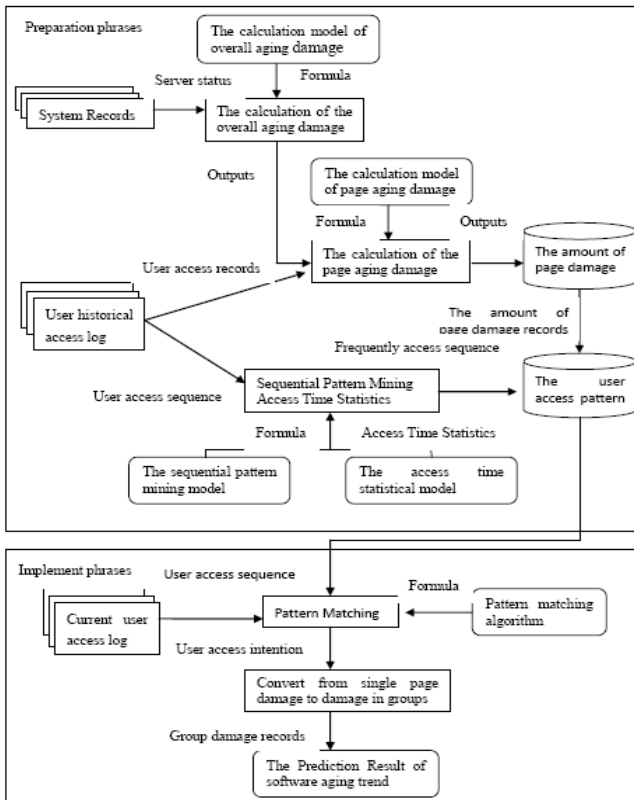


Fig. 1. The Prediction Process of Web Application System Aging Trend oriented to User Access Intention

tasks are to: remove noisy data, change the exiting data into usable formats, extract the useful formats and collect the multiple data sources. Data preprocess involve several step: data cleaning, user identification, session identification [9].

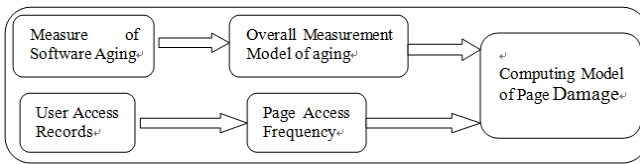
Data cleaning task is to get rid of some data records that do not reflect the user behavior in web log. So these “dirty data” should be removed by defining some rules. User identification is to identify each user accessing the system. The paper argues that different IP on behalf of different users; when the same IP, different operating systems or browsers on behalf of different users by default. Session identification is to identify the user access to a continuous page. Different users to access the pages belonging to different sessions, when time span of the user accesses two pages is greater than the definition of the Timeout value, which is considered to be two different sessions.

**The Amount of Page Damage**

This article assumes that each call to a single page, produce similar injury, and the amount of page damage to is defined as  $\phi$ . It is difficult to calculate the amount of damage  $\phi$  to each page directly, but the overall aging of the web server is the results of the role of each page, regard the overall wear and tear of software aging during the web server running as the dependent variable, the amount of damage to each page as an argument, the visits of each page as the coefficients of independent variables, then obtain the amount of damage to each page through multiple linear regression method [10] indirectly. Figure2 is the process of seeking the amount of page damage.

**User Access Sequential Pattern Mining Based on PrefixSpan Algorithm**

Definition 1. In user access pattern library, the sequence collection  $U=\{u_1, u_2, \dots, u_i, \dots, u_n\}$ , which  $u_i=\{p_i, t_i\}$ , where  $P_i$  refers to user access patterns,  $T_i$  refers to the mean residence time sequences of a page.



**Fig. 2.** The Calculation Process of Page Aging Damage

Definition 2. User access patterns  $P_i=\{x_1, x_2, \dots, x_k, \dots, x_n\}$ ,  $1 \leq k \leq n$ , which  $x_k$  refers to the page label.

It is not random for users to visit the web site’s pages, because there is a certain correlation among pages and the user’s access also has some models, these models can be mined by PrefixSpan algorithm. Based on PrefixSpan algorithm [12], steps of the user access pattern  $P_i$  as follows: Finding the sequence patterns sequence, with the length of one, in user access sequence collection, dividing the search space Sub according the elements in Seq. Recursive mining the subset of sequential pattern Sub.

Definition 3. The mean residence time sequences of a page  $T_i=\{t_1, t_2, \dots, t_{n-1}\}, 1 \leq k \leq n-1$ , which  $t_k$  refers to the mean residence time from page  $x_k$  to  $x_{k+1}$ .

When the users access a page, the residence time is different because the degree of interest, the user’s operating proficiency, reading speed are different. Some data may have significant differences with other data, the need to remove the noise data. After the noise removal, the mean residence time of a page is:

$$t_k = \frac{\text{Sum}(x_k, x_{k+1})}{\text{Count}(x_k, x_{k+1})} \tag{1}$$

Sum( $x_k, x_{k+1}$ ) refers to the total time from page  $x_k$  to page  $x_{k+1}$  in all users access record after the noise removal, Count( $x_k, x_{k+1}$ ) refers to the number of times of containing a sequence ( $x_k, x_{k+1}$ ) in all users record.

## 4 The User Access Intention-Oriented Prediction of Software Aging Based on Pattern Matching Algorithm

### 4.1 The Prediction of User Access Sequence Based on KMP Pattern Matching Algorithm

KMP algorithm is a classical pattern matching algorithm, its core idea is the process which use part of known match information to match the later information. This paper uses the KMP pattern matching algorithm to predict user access intention. The algorithm includes the follows steps:

- (1) Before carrying out pattern matching, change the pattern and text into the reverse form, that means compare the pattern with text forward one by one from the last character.
- (2) In matching, do not find a complete subsequence of the patterns in text, find a match sequence as long as possible. Regard the longest match sequence found from each text as the matching of the pattern subsequence under the text, marked as  $Pl_T(i)$ .

The returning results are  $Pl_T(i)$  and the predicted follow-up visits sequence. For example:

A record in users access pattern library is T : 1 3 4 6 4 8 1 2  
 The current users access sequence is P : 5 4 8

In the current access sequence, the last visited page class 8 has the greatest impact to the follow-up prediction result, and then followed by 4 and 5. So before applying the KMP algorithm set the current access sequence and the library record reverse. The results are as follows:

A record in users access pattern library is T : 2 1 8 4 6 4 3 1  
 The current users access sequence is P : 8 4 5

T' does not exist 5, so cannot find P' string in T', it is considered that the matching is unsuccessful in general KMP algorithm. However, in the improved KMP algorithm, find the longest matching string as much as possible if there is not a complete match, the record 5 minimal impact to the user intention in p', so remove the 5, and continue to find 8 and 4 in T'. There is a sting involving 8 and 4, so the match is successful. The general KMP algorithm returns the potion 2, where 8 4 located in T', while the improved KMP algorithm needs to return the follow-up sequence and matching: 2 1 2, which 2 1 is the reverse order sequence of the predicted follow-up visits, the last 2 means the length of matching is 2. Comparing the sequence T and P, the string 4 8 in P matches the string 4 8 in T, then it shows that this user probably accesses as this pattern, so we can predict this user's follow-up visit sequence is 1 2 according to the follow-up sequence is 1 2 in T.

**4.2 The Prediction of Software Aging**

Definition 1. Assume  $s=\{y_1,y_2,\dots,y_m\}$  indicate the current user access sequence, when  $y_1=x_p, y_2=x_{p+1}, \dots, y_q=x_{q+\sigma}$ , it means that the current user access sequence  $y_1$  to  $y_q$  matches with a record  $x_p$  to  $x_{q+\sigma}$  in the sequence pattern library, the  $\sigma$  means the number of the successful match items, called matching.  $\zeta=\{x_{p+\sigma},x_{p+\sigma+1},\dots,x_n\}$  is the user's follow-up visit sequence when the matching is  $\sigma$ .

The specific steps to predict web server aging are the server generates the current users access collection S based on the users access record. and matching any s belonging to S with the users pattern records in the users access pattern library to get several predictable sequence:

$$\begin{pmatrix} \sigma_1, n_1, \xi_1 \\ \sigma_2, n_2, \xi_2 \\ \vdots \\ \sigma_k, n_k, \xi_k \end{pmatrix} \tag{2}$$

Where  $\sigma$  is matching degree,  $n$  is the number of occurrences of each predictable,  $\zeta$  is predicted the follow-up visit sequence. Each of these sequences may be accessed, the bigger is the  $\sigma$  and the  $n$ , the greater the probability of access, so the access probability  $w_i$  is:

$$w_i = \frac{\sigma_i}{\sum_i \sigma_i} \times \frac{n_i}{\sum_i n_i} \tag{3}$$

Assumes that predicted sequence  $\zeta$  corresponding to the page damage is  $\Phi$ , then at time T, the individual user's expectations of the amount of damage is  $\Phi$ :

$$\Phi = \begin{cases} 0, \text{No access any page at time T} \\ \sum_{i=1}^k \omega_i \times \phi_i(T), \text{Others} \end{cases} \tag{4}$$

At time T, the expectations of web server aging is  $Jl$  :  $Jl=\Sigma\Phi$   (5)

## 5 Design and Implementation of Verification Experiments

### 5.1 The Prediction of User Access Intention

Divide the processed web log data into two categories: experimental data set and test data set. Firstly, mine the user access pattern library according PrefixSpan from experimental data set, then regard the first half of each sequence in test data set as the current user access sequence, and use match algorithm to predict the follow-up visit sequence, at last, compare the predicted result with the last part of test data set in order to verify the accuracy of the predictions.

Experiments were intercepted three groups of different sizes of experimental data, experiment results are shown in Table 1:

**Table 1.** Sense User Intention Experimental Result

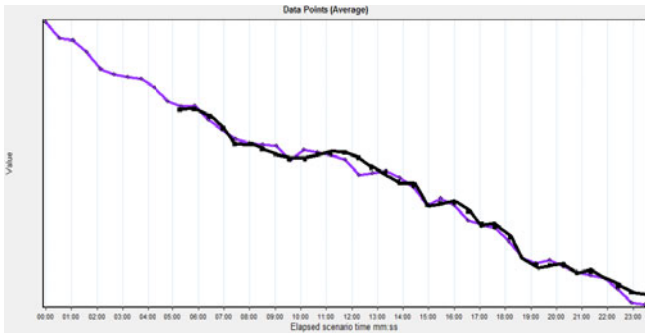
Experiments		1	2	3
Experimental Data set	Size(KB)	15	30	40
	Number of record	1200	2400	3230
Test Data set	Size(KB)	10	20	20
	Number of record	810	1600	1600
Number of frequent records		150	1770	2170
Accuracy		0.21	0.54	0.58

From the experimental results can be seen, with the experimental data set size increases, the number of generated frequently record and the accuracy of the prediction of user follow-up access intention also increase. When the experimental data set is large enough, user follow-up access sequence can be predicted accurately.

### 5.2 The Aging Prediction of Web Application System

Select 120 pages to stimulate the software aging process, when these pages are called, they take some memory but not release. Using loadrunner to stimulate users to call page continuously, then making prediction according the user-based web software aging method proposed in this paper, and compared with the actual value of software aging, as shown in Figure 3.

As can be seen from Figure 3, the software aging trend predicted by the user access intention-oriented prediction method for software aging trend has a high degree of fitting with the actual aging process, which prove that the method is valid and accurate.



**Fig. 3.** Experimental Results of Predicting the Software Aging Trend

## 6 Conclusion

Based on the past prediction research of software aging trend and considering more about the affect of user access load to web server aging, this paper proposed a user access intention-oriented prediction method. This method includes two main tasks, mining the user access intention and predicting the aging trend of server.

On one hand, by using sequence pattern mining algorithm, mine the access log records' pattern of web server, find the user access sequence pattern and further mine the frequent access sequence. On the other hand, by using matching algorithm, match the use access records with the access sequence in the pattern library; predict the user access intention and the server aging trend according to some information such as the amount of page damage. Finally, this paper identified the feasibility and accuracy of the user access intention-oriented software aging prediction method by using actual data; the experimental results show the highly fitting with the real server aging trend. In the next research, we will work to calculate a formula of the overall aging trend by the predicted user access intention, not just judged by the measurements.

**Acknowledgments.** We would like to thank the support of the project “the perception and evolution technology of network configuration software environment based on hierarchical adaptive strategies” (2009AA01Z122).

## References

1. Castelli, V., Harper, R.E., Heidelberger, P., et al.: Proactive Management of SoftwareAging. *IBM Journal of Research & Development* 45(2), 311–332 (2001)
2. Vaidyanthan, K., Trivedi, K.S.: Extended Classification of software Faults Based on Aging. In: *The IEEE Int'l. Symp. on Software Reliability Engineering, ISSURE 2001, Hong Kong* (2001)
3. Huang, Y., Kintala, C., Kolettis, N.: Software rejuvenation: analysis, module and applications. In: *Proc of FTCS225, Pasadena,CA* (1995)
4. Castelli, V., Harper, R.E., Heidelberger, P.: Proactive management of software Aging. *IBM JRD* 45, 311–332 (2001)

5. Huang, Y., Kintala, C., Kolettis, N., et al.: Software Rejuvenation: Analysis, Module and Applications. In: IEEE Intl. Symposium on Fault Tolerant Computing, FTCS 25
6. Alonso, J., Torres, J., et al.: Adaptive on-line software aging prediction based on machine learning. In: IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 765–771 (2010)
7. Garg, S., van Moorsel, A., Vaidyanathan, K., Trivedi, K.: A Methodology for Detection and Estimation of Software Aging. In: Proc. Ninth Int'l Symp. on Software Reliability, Eng., november 1998, pp. 282–292 (1998)
8. Yu, X., Guo, S., Peng, Z.: Web Log Mining based on Website Topic. In: The Seventh Wuhan International E-Business Conference, pp. 467–472 (2008)
9. Li, Z., Wan, Q.L., Xu, M.W.: Research on Software Rejuvenation. *Journal of Computer Science* 30(8), 1245–1252 (2003)
10. Yang, H.: Application of mathematical statistic, pp. 134–135. Tsinghua Publication, Beijing (2006)

# The Dynamic Control of the Process Parameters of MAG

Chuntian Li, Changhua Du, Huibin Xu, and Yi Luo

College of Material Science and Engineering, Chongqing University of Technology,  
No.69 HongGuang Road, Ba'nan District, Chongqing, P.R. China, 400054  
{lct0214, duchanghua, hbxu, luoyi}@cqut.edu.cn

**Abstract.** Based on the comprehensive analysis of welding circuit for MAG (Metal Active-Gas welding) arc power, a system model, through power, cable, wire, droplet and arc, is found, and a voltage function of wire is also developed according to distributed temperature function about a variable of GLW (Genuine Length of Wire). Then, a dynamic and multicycle droplet-voltage mathematic model is also found by the TWC (time-weight coefficient)  $\phi$  for the equivalent length of transferred droplet and by the transition of wire volume from a solid to a liquid state. Besides these, a arc-voltage function, involved the welding current, the GLW and the height of welding nozzle, is established logically and especially, Meanwhile, the correlative parameters is experimentally analyzed. At last, there made some acute and good dynamic control models of the process parameters of MAG on series of experiments in the paper.

**Keywords:** MAG, Droplet Transition, Parameter Model, Experimental analysis.

## 1 Introduction

In the MAG process, the self-regulation of arc play a role, but the parameters often fluctuates, such as arc current, arc voltage, they affect the stability of arc burning, thereby ,affect the quality of the joints. so, a few welding technologists have some research and effective way, and in the present, some corresponding control model have been found, there are: ①the regulating feed speed of wire based on voltage feedback model, [1] ②the vague control model of using current feedback [2] ③the adjustment model on the welding arc length for its highly stability, [3] ④ the arc voltage model by changing the supplied characteristics of the welding power, [4] these methods have all focused on the single factor or local links in MAG process, they ignore the effect that the droplet and the GLW on arc voltage, or consider changes only in the local part of the welding process but impact in the remaining part of the overall system, Obviously exists a great limitation.

Therefore, the paper takes MAG process including all internal and external loads (power-cables- the GLW-droplet-arc) as a whole, which is an interrelated and mutually constraining unity. Based on the GLW, it is built five mathematical models



about the MAG parameters relationship, including "power-cables-the GLW-droplet-arc" system model, the sub-model of length of welding arc and height of welding nozzle, the three sub-models of required voltage respectively from the GLW, droplet and arc. These models can promote the manufacture technology for MAG power, and can perform dynamic control of welding technology parameters. And can also provide other engineering practice and research to some theoretical guidance. The model, which is adopted in the paper, is finally from series of experiments.

## 2 The Dynamic Control Models of MAG Process

### 2.1 The System Control Model of MAG Process

In the study of the dynamic forming model, the addition from the end of solid welding wire to the front of welding electrode is defined as the GLW, the distance between the front of the semi-solid wire and liquid droplet is uniformly defined as the total droplet length. Based on the principle of Theremin equivalent and the law of Kirchhoff voltage, the system model is shown in Figure 1.

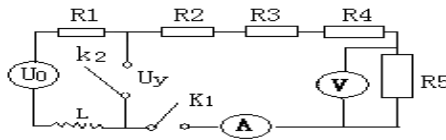


Fig. 1. Equivalent circuit sketch for the system model of welding process

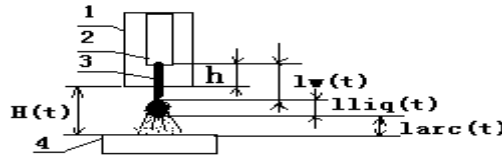
$$U_0 - L \cdot dI(t)/dt - U_{R1}(t) = U_{R2}(t) + U_{R3}(t) + U_{R4}(t) + U_{R5}(t) \tag{1}$$

$$U_{R2}(t) = I(t) \cdot \rho_2 \cdot l_1 / S_1 \tag{2}$$

In the above formula,  $U_0$  is the power voltage supply for all loads,  $L$  is the equivalent series inductance in welding power,  $I(t)$  for the welding current;  $U_y$  stands for the output voltage of the power;  $R_1, R_2, R_3(t), R_4(t), R_5(t)$  respectively stands for the static resistance of power, the static resistance of welding circuit cable, the dynamic resistance of the GLW, the equivalent dynamic resistance of droplet and arc,  $U_{R1}(t), U_{R2}(t), U_{R3}(t), U_{R4}(t), U_{R5}(t)$  respectively stand for the corresponding required voltage,  $\rho_2$  is resistivity of copper wires for the cable,  $L_1$  is the loop length of welding copper wire,  $S_1$  is for the cross-sectional area of copper wire.

### The Control Sub-model of Welding Arc Length and the Height of Welding Nozzle

During the welding process, electrode, nozzle, welding wire and the spatial distribution of arc shown in Figure 2:



1. Nozzle, 2. Electrode, 3. Wire, 4. Work pieces.

Fig. 2. The Equivalent structure diagram of electrode and arc

$$H(t) + h - l_w(t) = l_{liq}(t) + l_{arc}(t) \tag{3}$$

$$dl_w(t) / dt = V_s - V_m \tag{4}$$

$H(t)$ : welding nozzle height.  $h$ : the length from the level front of welding nozzle to the level of the electrode end, it is generally constant.  $l_w(t)$ : the GLW.  $l_{arc}(t)$ : arc length.  $l_{liq}(t)$ : droplet length,  $V_s$ : the feed speed of welding wire,  $V_m$ : wire melting rate.

**The Control Sub-model of Required Voltage from the GLW**

According to Ohm's law and resistance formulas:

$$U_{sol}(t) = U_{R3}(t) = I(t) \cdot R_3(T) \tag{5}$$

$$R_3(T) = \rho_{sol}(T) \cdot l_w(t) / S_w \tag{6}$$

Here,  $S_w$ : the cross-sectional area for wire.  $\rho_{sol}(T)$ : resistivity of wire, the relationship between it and temperature  $T$  generally can follow:

$$\rho_{sol}(T) = a \cdot T(y) + b \tag{7}$$

( $a, b$  is constant) The temperature distribution on the GLW can be approximated with a quadratic curve [5] expressed as:

$$T(y) = C_1 \cdot y^2 + C_2 \cdot y + C_3 \tag{8}$$

( $C_i, i = 1, 2, 3$ , for the constant,  $y$ : the distance from the front of electrode to somewhere on the direction of the wire.)

From (7), (8), formula (9) can be concluded:

$$\rho_{sol}(T) = \frac{1}{l_w(t)} \cdot \int_0^{l_w(t)} \{ a \cdot (C_1 \cdot y^2 + C_2 \cdot y + C_3) + b \} \cdot dy \tag{9}$$

By (5), (6), (9) known formula (10):

$$U_{sol}(t) = I(t) \cdot a / s_w \cdot \left\{ \frac{1}{3} \cdot C_1 \cdot l_w^3(t) + \frac{1}{2} \cdot C_2 \cdot l_w^2(t) + \left( C_3 + \frac{b}{a} \right) \cdot l_w(t) \right\} \tag{10}$$

**The Control Sub-model of Required Voltage for the Droplet**

In actual welding process, on the one hand, from the view of volume for melting drop, it can be measured in two dimensions: length of instant droplet and the dynamic radius of droplet. The volume depends mainly on the melting rate relative to temperature, electro-magnetic force, and the other. The size of droplet depends on the surface tension, gravity, arc thrust, electro-magnetic contraction force, the spot force, etc. [6] On the direction of feeding wire, the droplet's length  $l_{liq}(t)$  increase, the droplet's radius to decrease, on the contrary, the droplet radius to increase. Thus, based on volume and size of the droplet, from the point of integral perspective, the  $l_{liq}(t)$  is closely related to the melting rate and droplet's radius [7,8], it can be seen as a function of them, namely:  $l_{liq}(V_m, r_{liq}, t)$ . Here,  $V_m$ : melting rate,  $r_{liq}$ : the droplet's radius on the direction perpendicular to the wire's feed speed,  $t$ : the time for drop formation and melting transition, thereby, the formula (3) turns into (11).

On the other hand, when the droplet grow up, the arc space is objectively reduced, the total growth in length of the droplet is equal to the reduction of arc length, so that, if arc length  $l_{arc}(t)$  can be expressed as  $l_{arc}(V_m, r_{liq}, t)$ , the relationship between them such as the formula (12) as follows:

$$H(t) + h - l_w(t) = l_{liq}(v_m, r_{liq}, t) + l_{arc}(v_m, r_{liq}, t) \tag{11}$$

$$\frac{\partial l_{liq}(v_m, r_{liq}, t)}{\partial t} = - \cdot \frac{\partial l_{arc}(v_m, r_{liq}, t)}{\partial t} \tag{12}$$

Thus, in the cycle ( $T_{liq}$ ) of metal transfer, by the formula (11) and (12), to known the formula (13), (14):

$$\bar{L}_{liq}(T_{liq}) / \bar{L}_{arc}(T_{liq}) = \varphi / (1 - \varphi) \tag{13}$$

$$\bar{L}_{liq}(T_{liq}) = \varphi \cdot \{ H(t) + h - l_w(t) \} \tag{14}$$

Here:  $T_{liq}$ : Droplet "formation-growth- transition "period,  $t_1$ :the time for droplet formation , $t_2$ :the time for droplet down, that  $T_{liq} = t_1 + t_2$ ,  $\bar{L}_{liq}(T_{liq})$ : the time's weighted length during the  $T_{liq}$  for the droplet's form, growth and transition,  $\bar{L}_{arc}(T_{liq})$ : the time's weighted length for arc in  $T_{liq}$  ,  $\varphi$ : the coefficient for time's weighted length of droplet,  $\varphi=t_1/T_{liq}$  (if the droplet's transition for short-circuit form,  $\varphi=1/2$ ,if it for jet flow form,  $\varphi \leq 1/2$ ).So there the formula (15) set up:

$$R_4(ti) = \frac{\rho_{liq}(t) \cdot \bar{L}_{liq}(T_{liq})}{S_{liq}(T_{liq})} = \frac{\rho_{liq}(t) \cdot \bar{L}_{liq}^2(T_{liq})}{V_{liq}(ti)} \tag{15}$$

Where:  $\rho_{liq}(t_i)$ :the resistivity for droplet;  $S_{liq}(T_{liq})$ : the droplet's equivalent cross-sectional area corresponding  $\bar{L}_{liq}$  within the  $T_{liq}$ ,  $t_i$ : the transition period for the droplet  $i$ ,  $V_{liq}(t_i)$ : the volume of liquid droplet  $i$  in the  $t_{i-1}$ .

And since the formula (16), (17)

$$U_{liq}(t_i) = U_{R4}(t) = I(t) \cdot R_4(t_i) \tag{16}$$

$$V_{liq}(t_i) = \beta \cdot V_{sol}(t_i) \tag{17}$$

Where:  $V_{liq}(t_i)$ : the required voltage for droplet  $i$ ,  $R_4(t_i)$ : the equivalent resistance for droplet  $i$ ,  $V_{sol}(t_i)$ : the solid volume of droplet  $i$  in the  $t_{i-1}$ ,  $\beta$ : The blow-up coefficient for the wire's volume translation from solid to liquid. ( $1 < \beta$ )

Also because (18), (19) :

$$V_{sol}(t_i) = \int_0^{t_{i-1}} \pi \cdot r_w^2 \cdot v_m(t_i) dt \tag{18}$$

$$v_m(t_i) = l / t_i \tag{19}$$

According to the laws of heat capacity and the laws of energy, the equation (19) is known:

$$v_m(t_i) = \frac{\eta_1 U_{arc}(t_i) \cdot I_{arc}(t_i) + \eta_2 U_{sol}(t_i) \cdot I_{sol}(t_i)}{K_0 \cdot \Delta T} \tag{20}$$

$V_m(t_i)$ : the wire's melting rate,  $K_0$ : a constant,  $\Delta T$ : the temperature difference for wire melting before and after,  $\eta_1$ : the thermal efficiency coefficient from arc's heat,  $\eta_2$ : the thermal efficiency coefficient from resistance's heat.

Associated to (15), (16), (17), (18), (20), it have formulal(21):

$$U_{liq}(t_i) = \frac{I(t) \cdot \rho_{liq}(t) \cdot L_{liq}(T_{liq}) \cdot k_0 \cdot \Delta T}{\pi \cdot r_w^2 \cdot k_1 \cdot \beta \cdot t_{i-1} \cdot k_1 \cdot [\eta_1 \cdot I_{arc}(t_i) \cdot U_{arc}(t_i) + \eta_2 \cdot I_{sol}(t_i) \cdot U_{sol}(t_i)]} \tag{21}$$

Where:  $r_w$  as the wire's diameter,  $k_1$  for the integral coefficient. If the instant current changes is neglect able, in each cycle for the metal transfer, the output current  $I(t)$  from welding power,  $I_{arc}(t_i)$ , and  $I_{liq}(t_i)$  are synchronous changes in the establishment of equal time, that holds equation (22) in  $T_{liq}$ .

$$I(t) = I_{arc}(t_i) = I_{liq}(t_i) \tag{22}$$

Associated to (14) 、 (21) 、 (22), it has formula (23),

$$U_{liq}(t_i) = \frac{\varphi^2 \cdot \rho_{liq}(t) \cdot \{H(t) + h - l_w(t)\}^2 \cdot k_0 \cdot \Delta T}{\pi \cdot r_w^2 \cdot k_1 \cdot \beta \cdot t_{i-1} \cdot k_1 \cdot [\eta_1 \cdot U_{arc}(t_i) + \eta_2 \cdot U_{sol}(t_i)]} \tag{23}$$

From formula (23), it is known that if the other parameters remaining unchanged, the voltage for droplet is no direct relationship with the arc voltage, but is a reverse change with the arc voltage and genuine wire length, a positive change with the height of welding nozzle.

**The Control Sub-model of Required Voltage for Arc**

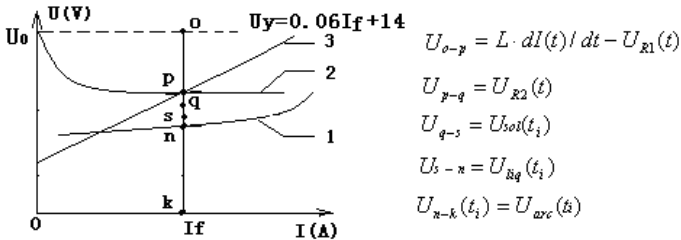
By the formula (1), (10), (23) to know:

$$U - L \cdot dI(t)/dt - U_{rl}(t) - U_{ke}(t) = U_{lr}(t) + U_{so}(t) + \frac{p_0 \cdot \{H(t) + h - l_w(t)\}^2}{p_1 U_{lr}(t) + p_2 U_{so}(t)} \tag{24}$$

Here :  $p_0 = \varphi \cdot \rho_{liq}(t) \cdot k_0 \cdot \Delta T$ ,  $p_1 = \pi \cdot r_w^2 \cdot k_1 \cdot \beta \cdot t_{i-1} \cdot \eta_1$ ,  $p_2 = \pi \cdot r_w^2 \cdot k_1 \cdot \beta \cdot t_{i-1} \cdot \eta_2$

The model, in the nature, it is the function of arc voltage, welding current, the genuine wire length and a height of welding nozzle, its establishment base on conditions: (1) not taking into account effect to the length of arc while droplet dropping down, (2) not to consider small spill from the pool surface to impact on the length of arc (3) the resistivity  $\rho_{liq}(t)$  for liquid droplet is constant, it don't change with temperature, and other physical parameters also unchanged, (4) in the time of short-circuit, the instantaneous arc current is equal to the welding current from the power output, (5) not to consider the MAG gas' impact on the translations of droplet.

The distribution of voltage corresponding to the model [8] is shown in Figure 3: when the welding current is  $I_f$ , the corresponding point of the power operating is at p, on the point n ,the arc can steadily combusting.  $U_0$  for the outputting voltage for no-load, the corresponding o-p section stands for the internal voltage required the power windings; the p-q section stands for the required voltage for the circuit cable; the q-s section means the shared voltage for the solid wire, the s-n section means the shared voltage for the droplet, the n-k section means the shared voltage for arc drop, namely:



1, the static characteristic curve for the welding arc, 2, the outputting characteristic curve for the power, 3, the exponential curve for the cable load from provisions.

Fig. 3. The welding graph of voltage distribution

The above analysis shows that to maintain stability of the welding parameters can be achieved by: (1) under fixing the outputting characteristics for the power, namely the curve 2 fixed, by appropriately adjusting the welding nozzle height, the arc current and arc voltage can achieve stability; (2) under fixing the outputting characteristics for the power, by adjusting the feed speed for wire to change the genuine wire length, that is, the feedback from the length of wire to the motors for the wire feed to ensure the stability of the arc current and the arc voltage, (3) to determine a good and fixed welding current, by matching the outputting voltage (curve 2 changes) and the height of welding nozzle (4) to determine a good and fixed welding current, by adjusting the outputting voltage (curve 2 changes) and the feed speed of wire simultaneously.

**The Experimental and Dynamic Analysis of the Control Model for MAG**

Using of relevant instruments and equipment, the paper has taken a specific parameterization on the model of the arc voltage [9], the experimental conditions are followed: the wire specifications are AWS ER70S-6, the wire diameter  $d=1.0\text{mm}$ ,  $\rho_{20}=0.018\Omega \cdot \text{mm}^2/\text{m}$  ( $20^\circ\text{C}$ ), the length of experimental wire is  $L_{exp}=13.5\text{cm}$ . By the

DC Resistance Tester, the resistance of  $L_{exp}$   $R_{exp1} = 54.4\Omega$  in the  $25^\circ\text{C}$ ,  $R_{exp2} = 58.3\Omega$  in the  $50^\circ\text{C}$ ,  $R_{exp3} = 59.34\Omega$  in the  $77.5^\circ\text{C}$ , so, according to formula (6), (7), the average of a, b is respectively  $7.36 \times 10^{-11}$ ,  $9.66 \times 10^{-9}$ . Then, to take a wire inside the beaker, the water of breaker is heated to  $40^\circ\text{C}$ , and measured random calibration points of wire temperature with an infrared temperature tester, the experiment principles are shown in figure 4, The results are  $y_{11} = 0\text{cm}$ ,  $T_{11} = 38.4^\circ\text{C}$ ,  $y_{12} = 2\text{cm}$ ,  $T_{12} = 31.9^\circ\text{C}$ ,  $Y_{13} = 8\text{cm}$ ,  $T_{13} = 28^\circ\text{C}$ , When the water is heated to  $60^\circ\text{C}$ ,  $y_{21} = 0\text{cm}$ ,  $T_{21} = 58.2^\circ\text{C}$ ,  $y_{22} = 6\text{cm}$ ,  $T_{22} = 29^\circ\text{C}$ ,  $Y_{23} = 15\text{cm}$ ,  $T_{23} = 28.3^\circ\text{C}$ , When the water is heated to  $70^\circ\text{C}$ ,  $y_{31} = 0\text{cm}$ ,  $T_{31} = 63.8^\circ\text{C}$ ,  $y_{32} = 5\text{cm}$ ,  $T_{32} = 30.2^\circ\text{C}$ ,  $Y_{33} = 8\text{cm}$ ,  $T_{33} = 28.5^\circ\text{C}$ , When the water is heated to  $80^\circ\text{C}$ ,  $y_{41} = 0\text{cm}$ ,  $T_{41} = 74^\circ\text{C}$ ,  $y_{42} = 4\text{cm}$ ,  $T_{42} = 36.7^\circ\text{C}$ ,  $Y_{43} = 14\text{cm}$ ,  $T_{43} = 27.2^\circ\text{C}$ . So, based on formula (8),

$$C_1 = 5.04C_3^2 - 3242.13C_3 + 524041.59$$

$$C_2 = -0.47C_3^2 + 286.56C_3 - 4418612$$

Then, the (8) becomes to the formula (25):

$$T(y) = C_3 + (5.04C_3^2 - 3242.13C_3 + 524041.59) \cdot y^2 + (-0.47C_3^2 + 286.56C_3 - 44186.12) \cdot y_3 \tag{25}$$

Then, the (10) becomes to the formula (26)

$$U_{sol}(t) = I(t) \cdot (9.4 \times 10^{-4}) \cdot \left\{ \frac{1}{3} \cdot (5.04C_3^2 - 3242.13C_3 + 524041.59) \cdot l^3 w(t) + \frac{1}{2} \cdot (-0.47C_3^2 + 286.56C_3 - 4418612) \cdot l^2 w(t) + (C_3 + 131) \cdot l w(t) \right\} \tag{26}$$

Based on the (24), (25), (26), it is known that: the Partial derivative of the arc voltage to welding current is such as the formula (27), the Partial derivative of the arc voltage to the GLW is such as the formula (28):

$$\frac{\partial U_{arc(t)}}{\partial I(t)} : -U/dt - R_1 - R_2 = \left( \frac{U_{arc(t)}}{I(t)} \right)' + \left( \frac{U_{sol(t)}}{I(t)} \right)' - \left( \frac{1}{p_1 U_{arc(t)} + p_2 U_{sol(t)}} \right)^2 \times p_0 [1.6 \times 10^{-3} - l_w(t)]^2 \times \left[ \left( \frac{p_1 U_{arc(t)}}{I(t)} \right)' + \left( \frac{p_2 U_{sol(t)}}{I(t)} \right)' \right] \tag{27}$$

$$\left( \frac{U_{sol}(t)}{I(t)} \right)' = 9.4 \times 10^{-4} \times \left[ \frac{1}{3} (5.04C_3^2 - 3242.13C_3 + 524041.59) l_w(t)^3 + \frac{1}{2} (-0.47C_3^2 - 286.56C_3 - 44186.12) l_w(t)^2 + (C_3 + 1.31 \times 10^3) l_w(t) \right]$$

Here:

$$\frac{\partial U_{arc}(t)}{\partial lw(t)} : -\left(\frac{-l \cdot d\dot{l}(t)}{dt lw(t)}\right)' - \left(\frac{U_R(t)}{lw(t)}\right)' - \left(\frac{U_{R_0}(t)}{lw(t)}\right)' = \left(\frac{U_{arc}(t)}{lw(t)}\right)' + \left(\frac{U_{sol}(t)}{lw(t)}\right)' \tag{28}$$

$$-\left(\frac{1}{p_1 U_{arc}(t) + p_2 U_{sol}(t)}\right)^2 \times 2p_0 [1.6 \times 10^2 - lw(t)] \times \left[ \left(\frac{p_1 U_{arc}(t)}{lw(t)}\right)' + \left(\frac{p_2 U_{sol}(t)}{lw(t)}\right)' \right]$$

$$\left(\frac{U_{sol}(t)}{lw(t)}\right)' = I(t) \times 9.4 \times 10^4 \times \left[ \frac{1}{3} (5.04C_3^2 - 324213C_3 + 52404159) l_w(t)^2 + \frac{1}{2} (-0.47C_3^2 - 28656C_3 - 441862) l_w(t) + (C_3 + 1.31 \times 10^2) \right]$$

Here:

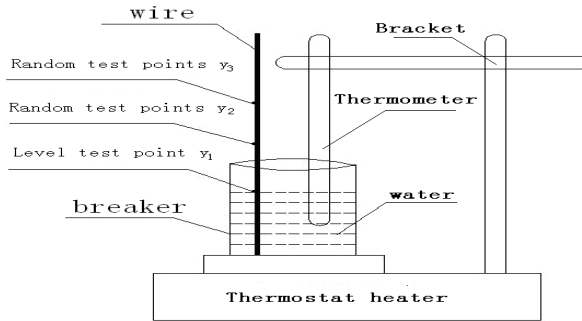


Fig. 4. The experiment principles of temperature distribution function for wire

### 3 Conclusions

(1) The control voltage model of the GLW are built, that is a cubic function of it, the control voltage model for the droplet derived a continuous multi-cycle transition, are also established.

(2) Through the system model, the function of arc voltage, welding current, GLW and welding nozzle height, are structured, it can provide a theoretical basis for the control of relevant parameters and for their stability in the MAG.

**Acknowledgments.** The paper is based on the high-tech project fund (No.2008EG115065) from The Ministry of Science and Technology of the People’s Republic of China, the author is very much appreciated.

This paper also attributed to National Nature Science Foundation of China (No.50975303), Thanks a lot again. acknowledgments here.

## References

1. Kang, M.J., Rhee, S.: Arc stability estimation and control for arc stabilization in short circuit transfer mode of CO<sub>2</sub> arc welding. *Science and Technology of welding and Joining* 6(2), 95–102 (2001)
2. Tusek, J.: (Inst za varilstvo Ptujška). Influence of wire extension length on the welding process. *International Journal for the Joining of Materials* 8(3), 112–120 (1996)
3. Zhu, Z., Bao, Y., Yu, S.: *Transactions of the China Welding Institution* 23(3), 26–28 (2002)
4. Mitat, T.: Automatic setting of the arc voltage using fuzzy logic. *Welding Review International*, 130–133 (1996)
5. Yin, S.: *The Principles and Adjustment of the Welding Equipments for MAG*, pp. 5–20, 40–44. Mechanical Industry Press, Beijing (2000)
6. Choisk, Yooed, K.: The dynamic analysis of metal transfer in Pulsed current gas metal arc welding. *J.Phys. D: Appl. Phys.* 31, 208–214 (1998)
7. Nemchinsky, V.A.: The rate of melting of the electrode during arc welding-the influence of discrete removal of the melt. *J. Phys. D: Appl. Phys.* 31, 1566–1570 (1998)
8. Shisheng, H.: The power for arc welding and digital control, pp. 35–36. Mechanical Industry Press, Beijing (2007)
9. Shisheng, H., Wensong, Y., Jiaxiang, X.: *Transactions of the China Welding Institution* 20(3), 20–24 (1999)



# Analysis of Occupational Accidents during Construction of Buildings Using Classification and Regression Tree

Chia-Wen Liao

Department of Civil Engineering, China University of Technology, Taipei, Taiwan

**Abstract.** There is a higher rate of occupational accidents during the construction of buildings than most other industries on average. However, steps can be taken to reduce worker risk through effective injury prevention strategies. In this article, classification and regression tree (CART) is employed in identifying the characteristics of occupational injuries in the construction industry. Accident reports during the period 2000 to 2009 are extracted from case reports of the Northern Region Inspection Office of the Council of Labor Affairs of Taiwan. The results show that CART can indicate some patterns of occupational injuries in the construction industry. Proposed inspection plans should be in accordance with the cause of injury and the high-risk subgroup. The findings identified in this article provide a direction for more effective inspection strategies and injury prevention programs.

**Keywords:** Building, construction, classification and regression tree (CART), occupational accidents.

## 1 Introduction

The construction of buildings is dynamic and hazardous due to the diverse and complex nature of work tasks, trades, and environment, as well as the temporary and transitory nature of construction workplaces and workforces [16]. Therefore, the risk of occupational accidents during the construction of buildings is far greater than in a manufacturing based industry [17-18].

Risk assessment is generally based on variable-centered statistical techniques that measure the relationship between a group of risk factors and occupational accident. Unless specified within the model, these methods do not examine the complex interactions among risk factors. Pattern-centered statistical techniques provide a different way to use risk assessments to determine risk classification. Pattern-centered statistical techniques identify subgroups within a sample that possess similar characteristics, allowing for an examination of constellations of risk factors within projects as opposed to identifying variables that predict risk across all projects. One type of pattern-centered statistical techniques is classification and regression tree (CART) analysis. CART is a nonparametric data mining technique that generates a decision tree used to classify a dataset based on a given outcome [14]. CART analysis selects the most important predictors and their interactions from a large number of variables generating unique constellations of variables that most accurately predict low- and high-risk groups.

This article examines the characteristics of occupational accidents during construction of buildings. Classification and regression tree (CART) is employed in evaluating the associations between different factors and in identifying the patterns of industrial occupational accidents. Accident analysis is used in 256 occupational fatalities to facilitate injury prevention strategy development.

## 2 Method

### Materials

Considering the differences in the characteristics of construction branches, the data related to building construction are extracted and analyzed in this article. We analyzed 256 accident reports of fatal occupational injuries during the period 2000 to 2009. All accident reports were extracted from case reports at the Northern Region Inspection Office of the Council of Labor Affairs of Taiwan. Relevant factors are derived from the accident reports. Each accident report was reviewed several times to itemize detailed information on each factor.

### Classification and Regression Tree

Classification and regression tree (CART) is a statistical procedure and primarily used as a classification tool [1, 3, 6-12, 14]. The CART methodology selects the most important predictive variables and their interactions, and generates unique constellations of target variables that most accurately predict low- and high-risk groups. Different types of predictive variables (categorical and continuous) can be integrated into a model. The machine-learning, probabilistic, non-parametric decision-tree method has been extensively exploited for vegetation mapping, ecological modeling, and remote sensing studies such as land use classification based on threshold values of various band data [4-5].

The CART methodology is especially attractive in data mining [1, 3, 6-10, 12]. First, due to their intuitive representation, the resulting model is easy to assimilate by humans [14]. Second, CART is a nonparametric data mining technique. Classification tree construction algorithms do not make any assumptions about the underlying distribution and are thus especially suited for exploratory knowledge discovery. Third, it uses an effective algorithm to cope with the missing data situation. It has been demonstrated that CART can still perform reasonably well when the missing data do not exceed 5%. Last, the accuracy of classification trees is comparable or superior to other classification models [2,4-5].

The CART methodology follows three steps. First, an overfitting tree is grown by recursive partitioning of the data. The second step called tree pruning, the sequence of nodes that should be eliminated based on a chosen evaluation measure to obtain a set of smaller trees is found. In the last step, an optimal tree is selected from the pruned trees [3-4].

CART grows trees only with binary splits. The principle is to recursively partition the target variable to maximize "purity" in the two child nodes. It systematically searches for the best variable out of all possible input variables that leads to the greatest improvement in the purity score of the resultant nodes. Several measures have been proposed for split evaluation: Gini, Twoing and Entropy. As the towing rule is not related to the impurity measure, and the Gini method is slightly better than

the Entropy tree fitting algorithm [15], we analyzed the Gini index only. Gini seeks for the largest category in the dataset and contends to isolate it from the other categories. The Gini impurity is then determined as:

$$i(t) = 1 - \sum_{j=1}^k (P_j(t))^2$$

Where  $i(t)$  is the impurity of node  $t$ ,  $P_j(t)$  is the portion of objects in node  $t$  that belongs to the  $j$ th class of the  $k$  classes present in the dataset [4, 13].

The method for tree pruning in CART is based on cost-complexity pruning, where both tree accuracy and complexity are considered [3, 4, 13]. The principle behind selecting the optimal tree is to find a tree with respect to a measure of misclassification cost on the testing dataset (or an independent dataset), so that the information in the learning dataset will not be overfit [4].

This instruction file for Word users (there is a separate instruction file for LaTeX users) may be used as a template. Kindly send the final and checked Word and PDF files of your paper to the Contact Volume Editor. This is usually one of the organizers of the conference. You should make sure that the Word and the PDF files are identical and correct and that only one version of your paper is sent. It is not possible to update files at a later stage. Please note that we do not need the printed paper.

We would like to draw your attention to the fact that it is not possible to modify a paper in any way, once it has been published. This applies to both the printed book and the online version of the publication. Every detail, including the order of the names of the authors, should be checked before the paper is sent to the Volume Editors.

### 3 Results

The CART analysis is shown in Figure 1. The optimal tree consists of 5 layers, namely “Whether the victim has work experience under 3 years”, “Whether the project is public”, “Whether the cause of injury is a fall”, “Whether the victim took no unsafe acts” and “Whether the victim is married or has children”. These are the 5 top-ranking factors in the database. Rare subgroups (those which occur fewer than 10 times, or which represent less than 1% of the data) are not treated in the present discussion.

#### Sample Characteristics

It can be seen from Tables 1 and 2 that most (89.45%) of the falls involve victims with less than 3 years work experience. Longer experience not only enables workers to build up professional skills, it also allows them to improve strategies for avoiding danger. Contractors typically require over 3 years of experience when hiring; this is why the factor was included in the present study. In general, the longer a worker’s experience, the less likely an accident is to happen: this trend emerges from Table 2. However, when work experience is over 20 years, the accident rate increases slightly. This may be because greater experience implies greater age and lesser physical strength, leading to a greater number of accidents.

From Tables 1 and 3 we can see that Falls constitute by far the most frequent form of accidents. In second and third place come Electric shocks and Collapses respectively. At 73.05% of the total, Fall accidents are the most common in the construction industry, more frequent than all other accident types taken together. The main reason for this is that it is difficult to take preventative measures, and when such measures are provided for, they are implemented only imperfectly.

From Table 1 we can further see that most of the projects are not public (government-funded) (85.94%). One possible reason is that most public projects are large-scale undertakings, and closer attention is paid to safety training and related facilities and procedures. 65.63% of victims did not take any unsafe actions, according to the statistical findings of this research, even though unsafe actions in themselves are very likely to lead to accidents. 58.98% of victims are married or have children: this may be because those with family responsibilities are more reliant on their jobs and salaries, and are more likely to take risks in order to protect their positions.

### Classification and Regression Tree Analyses

1. Analysis of the layer “Whether the victim has work experience under 3 years” and “Whether the project is public”:

It may be seen from Table 4 that those with more than 3 years work experience make up a majority (71%) of those suffering accidents on public projects. Conversely, an 88% majority of accident victims on non-public projects have less than 3 years’ experience. This may be because contractors on public projects recruit workers those with greater skills, owing to job requirements. These workers tend to have more than 3 years work experience, and this leads to a greater likelihood of accidents occurring in that group. Moreover, echoing Table 1, the proportion of accident cases on non-public projects amounts to 85.94%, as compared to 13.06% among the generally more experienced workers on public projects.

**Table 1.** Case characteristics (n=256)

Factor	n	%
Whether the victim has work experience under 3 years (EXP)		
Yes	229	89.45
No	27	10.55
Whether the project is public (PUB)		
Yes	36	13.06
No	220	85.94
Whether the cause of injury is fall (FAL )		
Yes	187	73.05
No	69	26.95
Whether the victim has no unsafe acts (SAF )		
Yes	88	34.38
No	168	65.63
Whether the victim is married or has children (MAR )		
Yes	151	58.98
No	105	41.02

**Table 2.** Distribution of work experience (n=256)

<b>Factor</b>	<b>n</b>	<b>%</b>
work experience		
Under 3 years	229	89.45
3-6 years	13	5.08
6-9 years	4	1.56
9-12 years	4	1.56
12-15 years	3	1.17
15-21 years	1	0.39
Over 21 years	2	0.78
<b>Total</b>	<b>256</b>	<b>100.00</b>

2. Analysis of the layer “Whether the cause of injury is a fall” and “Whether the project is public”:

Table 4 shows that the victims with more than 3 years experience (11%) and not working on public projects (29%) amount to only 9 data records, that is less than the requisite 10. All subgroups are treated as special cases which cannot be readily analyzed or explained, and is not discussed here.

Looking at the group of victims with less than 3 years’ experience, we can see that the proportion of victims suffering falls on public projects (61%) is lower than that on non-public projects (74%). This finding does not in fact demonstrate that “Whether the project is public” is a predictor of accident likelihood. A possible reason for this is that when less experienced laborers go on site, contractors take into account their lack of experience and assign them to tasks that are easier and safer, meaning that they suffer a lower rate of fall accidents.

**Table 3.** Distribution of Cause of injury (n=256)

<b>Factor</b>	<b>n</b>	<b>%</b>
Cause of injury		
Falls	187	73.05%
Electric shocks	25	9.77%
Collapses	15	5.86%
Falling objects	13	5.08%
Caught in between, and clamped	3	1.17%
Struck by and against	6	2.34%
falls on the same level	4	1.56%
Drowning	1	0.39%
explosions	1	0.39%
others	1	0.39%
<b>Total</b>	<b>256</b>	<b>100.00%</b>

3. Analysis of the layer “Whether the cause of injury is a fall” and “Whether the victim took no unsafe acts”:

As in the previous section, the number of victims with more than 3 years experience and not working on public projects only reached one or two data records, less than the

requisite 10. This subgroup was therefore eliminated from the data, and is not discussed here. Looking at other subgroups, we find that the incidence of fall accidents associated with unsafe behavior is lower than that of non-fall accidents with unsafe behavior. A possible explanation for this is that on sites where the risk of falls is high, because it is difficult to protect against accidents and safety levels are lower, workers are more likely to take dangerous actions, leading to a higher risk of accidents.

4. Analysis of the layer “Whether the victim took no unsafe acts” and “Whether the victim is married or has children”.

Table 1 shows that workers who are married or have children are more likely to suffer work accidents. Serial analysis shows, however, that this conclusion does not

**Table 4.** Optimal tree

EXP(N)=11%	PUB(N)=29%	FAL(N)=40%	SAF(N)=60%	MAR(N)=50%	0.16%
				MAR(Y)=50%	0.16%
			SAF(Y)=40%	MAR(N)=0%	0.00%
			MAR(Y)=100%	0.48%	
		FAL(Y)=60%	SAF(N)=27%	MAR(N)=17%	0.11%
				MAR(Y)=83%	0.54%
	SAF(Y)=73%		MAR(N)=33%	0.58%	
		MAR(Y)=67%	1.17%		
	PUB(Y)=71%	FAL(N)=12%	SAF(N)=0%	MAR(N)=0%	0.00%
				MAR(Y)=0%	0.00%
			SAF(Y)=100%	MAR(N)=0%	0.00%
			MAR(Y)=100%	0.94%	
FAL(Y)=88%		SAF(N)=29%	MAR(N)=50%	1.00%	
			MAR(Y)=50%	1.00%	
	SAF(Y)=71%	MAR(N)=20%	0.98%		
	MAR(Y)=80%	3.90%			
EXP(Y)=89%	PUB(N)=88%	FAL(N)=26%	SAF(N)=31%	MAR(N)=37%	2.34%
				MAR(Y)=63%	3.98%
			SAF(Y)=69%	MAR(N)=31%	4.36%
			MAR(Y)=69%	9.69%	
		FAL(Y)=74%	SAF(N)=35%	MAR(N)=42%	8.52%
				MAR(Y)=58%	11.77%
			SAF(Y)=65%	MAR(N)=51%	19.21%
			MAR(Y)=49%	18.46%	
		PUB(Y)=12%	FAL(N)=39%	SAF(N)=27%	MAR(N)=67%
	MAR(Y)=33%				0.37%
	SAF(Y)=73%			MAR(N)=38%	0.16%
			MAR(Y)=63%	1.92%	
	FAL(Y)=61%		SAF(N)=41%	MAR(N)=29%	0.77%
				MAR(Y)=71%	1.90%
		SAF(Y)=59%	MAR(N)=30%	1.15%	
		MAR(Y)=70%	2.69%		

apply in cases where the victim took no unsafe actions. The group of victims with less than 3 years' experience, not on a public project, whose accident type was a fall, who did not take any unsafe actions and was unmarried with no children constituted, at 19.21%, the largest subgroup in our data. This result shows that "Whether the victim is married or has children" may have no direct relationship with accident occurrence.

## 4 Conclusions

In this article, classification and regression tree (CART) is employed in identifying the characteristics of occupational injuries in the construction industry. The analysis reveals that CART can indicate some patterns of occupational injuries in the construction industry. In the high-risk environment of construction sites, it is not enough to rely on safety equipment. One must also depend on safety education and training, alerting laborers to be aware of what is going on around them. Only then can personal safety be safeguarded. Proposed inspection plans should be in accordance with the cause of injury and the high-risk subgroup. The findings identified in this article provide a direction for more effective inspection strategies and injury prevention programs.

## References

1. Zhong, M., Georgiopoulos, M., Anagnostopoulos, G.C.: A k-norm pruning algorithm for decision tree classifiers based on error rate estimation. *Mach. Learn.* 71, 55–88 (2008)
2. Gehrke, J., Ramakrishnan, R., Ganti, V.: RainForest-A Framework for Fast Decision Tree Construction of Large Datasets. In: *Data Mining and Knowledge Discovery*, vol. 4, pp. 127–162 (2000)
3. Suknovic, M., Delibasic, B., Jovanovic, M., Vukicevic, M., Becejski-Vujaklija, D., Obradovic, Z.: Reusable components in decision tree induction algorithms. *Comput Stat.* (in press)
4. Cheng, W., Zhang, X., Wang, K., Dai, X.: Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution. *Environ. Monit. Assess* 158, 419–431 (2009)
5. Sledjeski, E.M., Dierker, L.C., Brigham, R.: The Use of Risk Assessment to Predict Recurrent Maltreatment: A Classification and Regression Tree Analysis (CART). *Prev. Sci.* 9, 28–37 (2008)
6. Hödar, C., Assar, R., Colombres, M., et al.: Research article Genome-wide identification of new Wnt/ $\beta$ -catenin target genes in the human genome using CART method. *BMC Genomics* 11, 348 (2010)
7. Chayama, K., Hayes, C.N., Yoshioka, K., et al.: Factors predictive of sustained virological response following 72 weeks of combination therapy for genotype 1b hepatitis C. *J. Gastroenterol.* 46, 545–555 (2011)
8. Moon, T., Kim, Y., Kim, H., Choi, M., Kim, C.: Fuzzy rule-based inference of reasons for high effluent quality in municipal wastewater treatment plant. *Korean J. Chem. Eng.* 28, 817–824 (2011)
9. Ruberg, S.J., Chen, L., Stauffer, V.: Identification of early changes in specific symptoms that predict longer-term response to atypical antipsychotics in the treatment of patients with schizophrenia. *BMC Psychiatry* 11, 23 (2011)

10. Yu, L., Wang, L., Yu, J.: Identification of product definition patterns in mass customization using a learning-based hybrid approach. *Int. J. Adv. Manuf. Technol.* 38, 1061–1074 (2008)
11. Gaudart, J., Poudiougou, B., Ranque, S., Doumbo, O.: Oblique decision trees for spatial pattern detection: optimal algorithm and application to malaria risk. *BMC Medical Research Methodology* 5, 22 (2005)
12. Hong, W., Dong, L., Huang, Q., Wu, W., Wu, J., Wang, Y.: Prediction of Severe Acute Pancreatitis Using Classification and Regression Tree Analysis. *Dig. Dis. Sci.* (in press)
13. Pesch, R., Pehlke, H., Jerosch, K., Schröder, W., Schlüter, M.: Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. *Environ. Monit. Assess* 136, 313–325 (2008)
14. Breiman, L., Friedman, J.H., Stone, C.J., Olshen, R.A.: *Classification and regression trees*. Wadsworth, Belmont (1984)
15. Breiman, L.: Decision-tree forests. *Machine Learning* 45, 5–32 (2001)
16. Kines, P.: Construction workers' falls through roofs: fatal versus serious injuries. *Journal of Safety Research* 33, 195–208 (2002)
17. Larssona, T.J., Field, B.: The distribution of occupational injuries risks in the Victorian construction industry. *Safety Science* 40, 439–456 (2002)
18. Sawacha, E., Naoum, S., Fong, D.: Factors affecting safety performance on construction sites. *International Journal of Project Management* 17, 309–315 (1999)



# Weapons Program Based on an Integrated Management Model of the Hall

Yongtao Yu<sup>1</sup> and Ying Ding<sup>2</sup>

<sup>1</sup> Institute of Strategy and Policy Naval Academy of Armament, Beijing, China

<sup>2</sup> Beijing Sino-Ocean JIYE Property Management CO., LTD, Beijing, China

**Abstract.** With the progress of weapons programs, project objectives, planning, control and management of the main targets have also changed, such as information exchange projects in the participating units close enough to some extent, project management is characterized by a certain degree of non-continuous and mutual independence. The large-scale weapons program trend has led to a long-term development projects, including users of the participating units, including units in the project implementation process needs, functions and roles change. In order to maximize project management efficiency, to build a weapons system on the Hall project life-cycle concept of integrated management model, the completion of the project, including weapons, decision-making stage, design stage, production stage and equipment, process analysis phase of integrated management, forming a weapons project integration management organizational structure. Hall-based weapons system integration project life-cycle management to help solve a class of high-investment, high-interest, long-period large-scale project management issues.

**Keywords:** Hall systems, program management, integration.

## 1 Introduction

Weapons program and more for large investment, high interest, long-period large-scale project, its technical and economic uncertainties that many, and the complex internal and external environmental conditions, project management is characterized by a certain non-continuous and mutual independence. Therefore, only set up the system, complete weapons integration project life-cycle management, engineering practice in order to meet the urgent needs of the project management requirements and higher [1].

Currently, the industry on project management theory and applications of basic and key issues, such as the composition of the whole life cycle, management, and management tools and other objects, carried out extensive research and made a large number of results. At the same time, while scholars often the potential problems of project management gains, weapons project integration management researchers often do not pay sufficient attention. This leads to the Hall system life-cycle integration of project management implications have not been adequately studied.

This article attempts to study hall system in the weapons program management, weapons of a typical project cycle management as the object of study to build a conceptual model of integrated management of weapons programs, set up a management organization, analysis of management processes, and as a design and implementation weapons program based on integrated management information system.

This chapter is organized as follows: Section 2 outlines Hall systems theory. Section 3 to build a weapons system on the Hall project life-cycle concept of integrated management model, given the organizational structure of the integrated management of weapons programs, completed the project management process integration analysis in Section 4 weapons program of integrated management information system examples. Section 5 summarizes the full text.

## **2 Hall System Principle Outlined**

Hall, Three-dimensional structure, also known as Hall systems engineering, systems engineering experts in the United States Hall (A • D • Hall) in 1969 as a systems engineering methodology. It appears to solve large and complex system of planning, organization; management provides a uniform way of thinking [2]. Hall, Three-dimensional structure of the systems engineering activities throughout the convergence process is divided into before and after the close of the seven phases and seven steps, and also considered for the completion of these stages and steps required for a variety of professional knowledge and skills. Thus, the formation by the time dimension, and knowledge-dimensional logical dimension consisting of three-dimensional spatial structure [3]. By the time dimension, logic and knowledge-dimensional composition of the three-dimensional space-dimensional structure system hand vividly describes the system engineering framework. One the other hand, at any stage and every step can be further expanded to form a hierarchical tree system [4].

## **3 Research Weapons Integration Project Life-Cycle Management Model**

Weapons integration project life-cycle management model is defined by the user units (military-related decision-making bodies, etc.) led unit has overall responsibility for professional advice, from all the major units involved in the election of one or two experts were composed of one project life cycle management group, the whole life cycle of the main units involved, the management of content, the project management phases combine to achieve the organization, resources, objectives, responsibilities, benefits and other integration-related information between the participating units to communicate effectively and fully shared to the user units and other stakeholders to maximize the value of the project units to provide the product. Integrated management of the project life-cycle groups in the composition of weapons project life cycle, are not immutable, should be mainly involved with the phase adjustment unit changes, add new items in units of experts, to facilitate the achievement of overall project effective management[5].

### 3.1 Build a Conceptual Model of Integrated Management

Start of the study, the first three-dimensional structure model based on the Hall theory, a weapons project life-cycle management integration shown in Fig. 1. Weapons program life-cycle management model integration covers three main areas: integration of the participating units, elements of management integration, integration of the management process [6]. To achieve the integration of the participating units, the units will help to break the project time, scope and content of the work on the boundaries, to promote the integration and management elements of the management process integration. Management process to achieve integration, but also requires management to break the phase interface, the implementation of integrated management elements played a role in promoting. The management factor also in turn promotes the implementation of the integration process of integration. On this basis, operational processes, organizational structure and information platform is an integrated weapons program life-cycle management of the three basic elements.

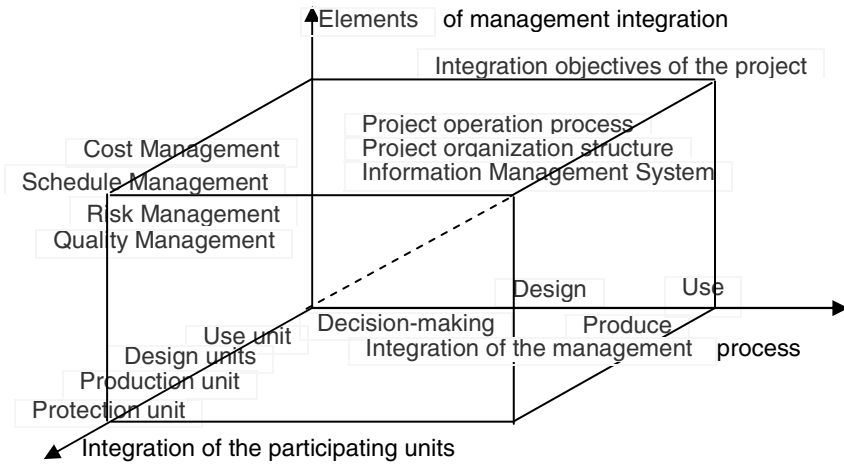


Fig. 1. Schematic diagram of an integrated management model weapons program

### 3.2 Design Integrated Management Process

Weapons and integrated management of project life-cycle mode operation of the project process flow with the traditional operation of the project has some similarities, but more emphasis on the balance of the project in the target unit, effective flow of information and the application of concurrent engineering. Operational flow of weapons project phases shown in Fig. 2.

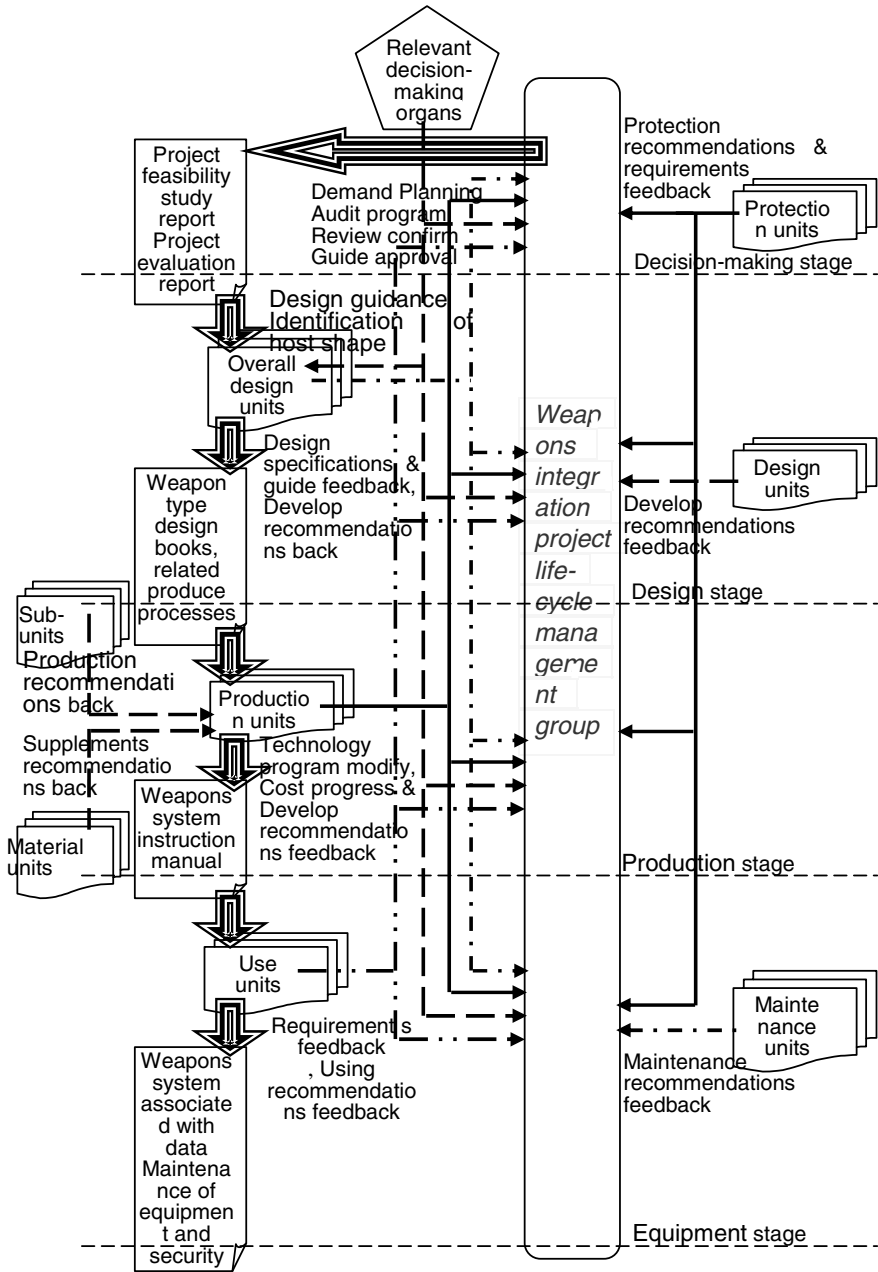


Fig. 2. Project stage Flow chart

Project Life Cycle Management Group as the main responsibility for integration and coordination unit, responsible for collecting information from all participating units, to determine the initial program and feedback to the relevant decision-making bodies (the military). Combat-related decision-making body considering the building, carry out mission requirements, force structure and the preparation of financial constraints and other conditions, to determine the optimal program. Project management team to refine the optimal solution demonstration, seek advice and design units, and timely analysis and collation of information of all kinds, and finally made weapons project feasibility study report and project evaluation reports.

Weapons design phase of the project operation process, the overall design of the system unit as the main responsible unit, to feasibility studies, conceptual design and planning requirements as the main design basis, project life cycle through integration with other management groups. The participating research units in repeated discussions on the design, determine the design that meets your needs, access to relevant decision-making bodies (the military) recognition and identification after completion of the design shape. The model design document and related materials, production technology transfer production assembly units.

Stage of the operational flow of weapons produced by the project, the main responsibility for production and assembly units for the coordinator to design the book and related types of weapons manufacturing materials as the main production basis, production process, considering the use and protection units. Use protection recommendations, feedback is given after repeated discussions execution. If the production process required to modify the design, via the integrated management group to submit changes to the design units in aggregate demand by the design unit to make design changes after production assembly unit in order to implement specific changes.

Stage of the operational flow of weapons and equipment items, the use of units (fielded forces, etc.) as the main responsible person in the first few stages of the project to collect data based on the use of weapon systems, combined with equipment used to protect and maintenance of a comprehensive assessment of the weapons program. And evaluation results back to the design unit, does not meet the requirements of items by the production unit to improve the final delivery of qualified weapon systems.

### **3.3 Analysis of Integrated Management Structure**

Weapons involved in the project management organization project management and responsibilities, authority and mutual relations division of a team arrangements and facilities, including user units, consulting units and other units involved in the project management work for the project management organization established[7]. Weapons program life-cycle management can be integrated organizational structure shown in Fig. 3. User units (military-related decision-making bodies, etc.) as the project's top decision-makers to oversee and manage the project life-cycle integrated management group, the project has the ultimate decision-making control over the final decision on project implementation and contract parties, while organization, leadership and supervision of the work to fulfill the user unit review, approval and authorization of the power to make the project goals. Project Life Cycle Management Group is an

integrated unit commissioned by the user phase of the project management of the main weapons, each unit is also responsible for the coordination of the various stages, the project participants through the project life-cycle units in the integrated management of groups involved in the project ahead of schedule, full participation to complete its task.

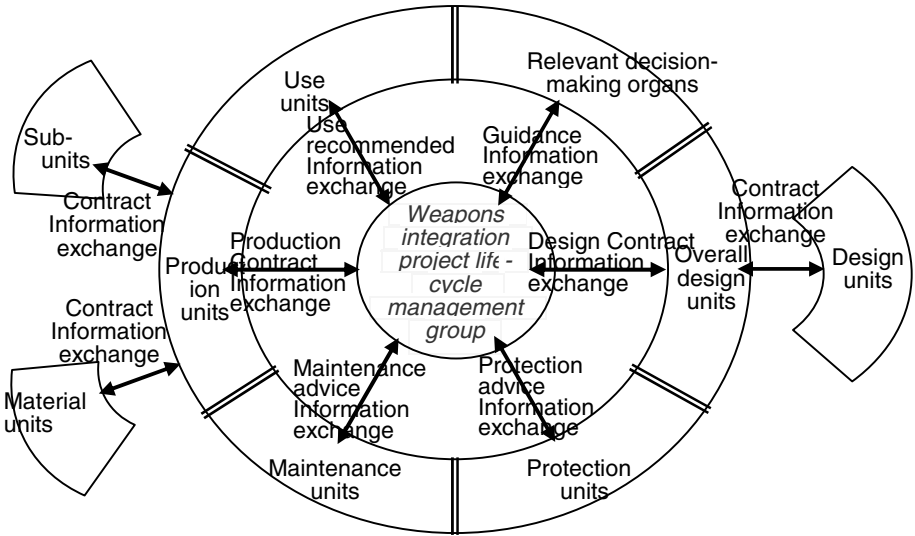


Fig. 3. Projects Organization Chart

#### 4 Build a Weapons Program Integrated Management Information System

Weapons to achieve integrated management of the project life-cycle model units, the various stages of effective communication and sharing of information, we must first achieve the information integration. Weapons program integrated information management system to ensure accurate and timely complete the project decision-makers to grasp fully reflect the implementation of the project information, and the participants of the project to achieve the flow of information and data sharing, to help project decision makers to make scientific decisions. Integrated information management system project structure design shown in Fig. 4.

The system can make use of computer and information processing technology, by building a unified central database and integrated information management platform project, the use of common, unified management of language and rules, the various stages of project implementation, including weapons, different functional blocks of data information integration, and the participants of the project to provide personalized information resources. Its implementation will help project the participating units to share information, exchange and transfer, making the project all participating units to communicate between the co-operations into an orderly, but also to ensure the project's centralized data management, digital, integrity and consistency.

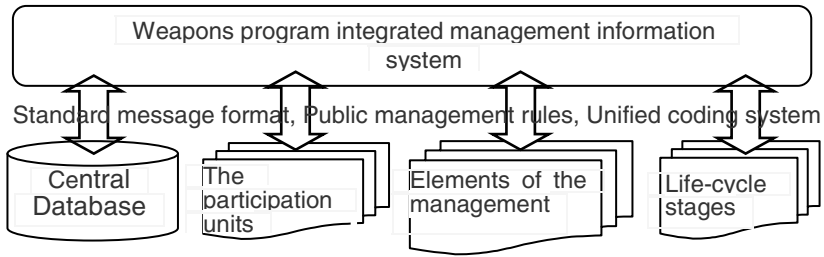


Fig. 4. Integrated information management system project structure

## 5 Conclusion

This reference life-cycle concept first proposed to build weapons based on the Hall system integration project life-cycle management concept model of the weapon for an objective analysis of the project management process; then the conclusions based on analysis of project management of weapons, weapons programs to form an integrated management structure, completion of the project, including weapons, the decision-making phase, design phase, production phase and the phase of the integrated management of equipment, process analysis, project running in arms during the whole life continued uninterrupted implementation of effective management; to achieve maximum benefit for the pursuit of weapons program .

Next, will continue to enrich, update and deepen the analysis of the participating units, weapons and items to complete project life-cycle integrated weapons management conceptual model, and the introduction of intelligent database research to improve the integrated management information systems weapons program efficiency.

## References

- [1] Bai, S.: Modern project management. Beijing Machinery Industry Press (2003)
- [2] Chen, H.: Project management. China Building Industry Press (June 2001)
- [3] Lv, Y.: Systems Engineering. Beijing Northern Jiaotong University Press (2003)
- [4] Qiu, W.-H.: Project management. Beijing Science Press (2001)
- [5] Tan, Y.: The system engineering principles. National Defense University Press (2003)
- [6] The institute of Internal Auditors, Applying COSO's Enterprise Risk Management—Integrated Framework (September 2004)
- [7] Committee of Sponsoring Organizations of the Treadway Commission, Enterprise Risk Management —Integrated Framework, Executive Summary (September 2004)

# Opinion Search and Opinion Mining: A New Type of Information Technology Bringing about New Scientific and Talent Revolution

Liu Yidong

Institute for the History of Natural Science, Chinese Academy of Sciences  
No.55 Zhongguancun East Road, Beijing, China 100190  
liuyidong.cn@hotmail.com

**Abstract.** The application of opinion retrieval and opinion mining in scientific research and academic appraisal would strongly promote the popularization of opinion search and conclusion search. They would significantly reduce research costs arising from repetitive researches and covert repetitive researches. Thereby, the output-input ratio of research costs would improve significantly, and the resource allocation would be optimized. Accordingly, experts of opinion retrieval and opinion mining need to conduct and prepare for related researches. The application of these new information technologies would also bring about a series of revolution in scientific research and academic appraisal, triggering scientific and talent revolution.

**Keywords:** Opinion retrieval, opinion mining, opinion search, conclusion search, repetitive research, research cost, scientific revolution, talent revolution.

Opinion retrieval is to search texts containing user's opinions, views or insights by searching keywords. When people enter the keywords, opinions and views related to the keywords will be retrieved. In recent years, opinion retrieval has become a key topic in information retrieval. Since 2006, in TREC evaluation, opinion retrieval has been used in blog-based information retrieval (Xu & Lin, 2010). Similar to opinion retrieval, opinion mining is another popular topic. Opinion mining is a method to automatically extract and analyze the subjective texts describing non-fact, which is a comprehensive research area across different disciplines. It is related to text mining, information extraction, information retrieval, machine learning, natural language processing, probability theory, statistical data analysis, ontology, corpus linguistics and visualization technology, etc. (Yao, et. al., 2008). Opinion retrieval and opinion mining are developing technologies. Currently, they only apply to commercial and social areas, such as e-commerce, business intelligence, market analysis, public comment, information monitoring, public opinion survey, digital learning, newspaper editing and corporate management, etc. For example, these technologies can fast obtain the sorted statistic results of comments on some online products (e.g. notebook computers) and provide the results to manufacturers. These results can help manufacturers to further improve their products, or provide references for prospective customers to select products, or help distributors to make orders (Yao et. al, 2008). Of course, these



applications are necessary. However, the application of opinion retrieval and opinion mining are overlooked in the scientific research area. Google Scholar is just a beginning. This paper analyzes the necessity, urgency and significance of the application of opinion retrieval and opinion mining in scientific research area from the perspectives of technology operation and academic appraisal. It can be concluded that the development and application of these new types of information technologies can bring science into “searching and analyzing age, online demonstration age and public supervision age” (Liu, 2008a). These technologies can efficiently enhance the output-input ratio of scientific research expenses and optimize the allocation of technology resource, and thus bringing about scientific and talent revolution. This paper mainly consists of three parts. First, it explains the significant implementation of such substantial searches as “opinion search / conclusion search”. Second, it confirms the role of academic priority system in promoting the science development as the driving role of patent system in the technology advancement. Third, it discusses how opinion retrieval and opinion mining shall improve and develop in order to meet the application requirements of scientific research and evaluation.

## **1 Implementing Opinion Search / Conclusion Search to Single Out Excellent Results and Excellent Talents**

The development of scientific research increasingly plays an important role in the modern society. The investment on scientific research is also increasing vastly. The government, enterprises, public, scientific and technical circles pay more and more attention to issues such as how to use the huge amount of scientific research expenses, improve output-input ratio and optimize the resource allocation. These issues are particularly significant in China. Whether we can improve the efficiency of scientific research expenses is a key to develop the scientific research and improve self-innovation capability of China. The author believes that the pointcut to improve the efficiency of scientific research expenses is to largely eliminate the repetitive researches and covert repetitive researches. These repetitive researches are costly and foster the academic corruption. Hence, really creative results and innovative scholars are overwhelmed in these repetitive researches. The researches that deserve to be studied are ignored, while the ones that shall be ignored are investigated heavily. Even the phenomenon of reverse elimination occurs. The author believes that the application of opinion search / conclusion search in academic research is a solution to this problem and the “academic priority system” shall be established. “Opinion search / conclusion search” is different from the general result search. It is also different from innovative key point search. The objects of opinion search / conclusion search are the key points or main research conclusions of academic results (sometimes, these two are the same). The key innovative points also refer to key points or main research conclusions. It is noticeable that the core feature of research innovation is that its key points or main research conclusions are innovative. If the research objects or research materials or research methods are innovative, but the key points or main research conclusions are not innovative, such research results are not innovative because they do not have innovative “output”. This situation can be regarded as “covert repetitive research”.

The author calls “opinion search / conclusion search” the “substantial search” of results. The reasons are explained as follows.

Strictly speaking, academic research studies the “academic issues” (including finding problems and solving problems). The “output” of academic research is the solution to investigated issues. Or, generally speaking, the core of academic results is research conclusions, including solutions and related views. The author defines the academic opinions as the conclusions made in academic researches, including solutions and related opinions. The definitions of academic opinions and research conclusions literally are the same. Sometimes, their contents are the same too. The difference between the two represents in the “related opinions” part. Generally, research conclusions include, but not limited to academic opinions, whose contents are more than academic opinions. Currently, the definition of innovative results has not been specified. Considering all innovative contents (apart from “output”) such as research objects, research materials and research methods as innovative results lowers the evaluation criteria. This leads to many ordinary researches being considered as innovative researches and a waste of research expenses.

The substantial search such as “opinion search / conclusion search” is objective and practicable. More importantly, the application of these technologies can immediately distinguish the disqualified results from qualified results (including excellent results). Standing out in the substantial search is a prerequisite for excellent results. This simple measure can avoid the repetitive researches and covert repetitive researches. It can also improve the output-input ratio of research expenses and single out the excellent research results and excellent talents. The main three reasons are listed as follows:

First, only new knowledge is valuable. In the information age, the broadcast and copy of knowledge is very easy. Comparatively, the knowledge production is much more difficult. Therefore, new knowledge production is much more valuable than repetition of existing knowledge (some blocked areas are exceptional. For in such cases, the broadcasting cost is high. So, repetitive production is valuable.). Second, the new knowledge production is not easy. And proposing new academic opinions is more difficult. Scientific researches produce knowledge through finding and solving problems. Not every sentence with a question mark is an academic problem. The academic problems are proposed under certain theoretical background and framework. The advancement of existing knowledge or creation of new knowledge is based on reviewing and summarizing researches on the same or similar academic problems. The objects to be studied as well as phenomena and issues to be analyzed are many. However, the academic topics to be specified are limited. Scientific researches summarize or extract new general academic opinions through solving a few academic problems. The global researchers follow the universe academic procedures and standards to conduct the researches. So the competition is fierce as researchers rush to publish the new opinions and new conclusions that are easily identified. Therefore, it's difficult to propose new opinions and new conclusions. The academic circle has been aware of this phenomenon. For example, Huang Ping, Researcher of Chinese Academy of Social Sciences points out that “it is very difficult to achieve a bit of result or make a breakthrough in academic researches” (Huang, 2002).

Third, qualified scholars are a world's championship in their segmented research field. Scholars who can produce reasonable new knowledge and propose new opinions are qualified ones. At least, they are the world's championship or the world's record

holder in their segmented research area. According to the academic standard and procedure, a research shall be advanced or explored based on reviewing literature related to the research problems. A researcher should at least have one point better than other scholars, or temporarily or partially leads in the same research area. In academic researches, there is only gold medal winner without silver medal winner. Therefore, the scientific development relies on talents. Talents that can conduct academic researches are excellent innovative ones. Excellent results and excellent talents can be singled out through identifying innovative and accurate opinions.

Obviously, the application and development of opinion retrieval and opinion mining can provide technology support for implementing substantial search such as “opinion search / conclusion search”.

## **2 Establishing Academic Priority System to Avoid Repetitive Researches and Improve Output-Input Ratio**

Repetitive researches and covert repetitive researches have wasted a large amount of research expenses. In order to reduce such waste, only implementing substantial search (opinion search / conclusion search) is not enough. The academic priority system is needed to provide a system guarantee. As we all know, the establishment of patent system plays an important role in advancing the technology development. However, an academic priority system similar to patent (technology priority) system has not been established so far. The confirmation and acquisition of academic priority has been in an inefficient, postponed and unclear situation for a long time. Academic priority system recognizes intellectual results, which is also known as academic priority recognition system. It offsets the inadequacy of current academic recognition, which can significantly reduce repetitive researches and covert repetitive researches. The important innovative research results can be used by other researchers as a basis to improve the efficiency of knowledge production and strongly promote knowledge growth and academic advancement. The academic priority system will learn from the patent system but different from it in response to the characteristics of academic researches, which includes the application, search and standard recognition of academic priority as well as the procedures of reporting and demonstration, databases and retrieval system of academic priority (Liu, 2007). We can use a system similar to the patent system to recognize the academic priority. Because scholars who can obtain the academic priority (i.e. academic innovative insights) in the key or popular academic topics are a few, we only need to list the academic priority and we will know the level of scholars. It is like inventors who know the level and capability of other inventors through displaying their invention patents.

The establishment of academic priority system will facilitate the communication and cooperation between scholars. As the patent system promotes the technology advancement, the establishment of academic priority system will inevitably promote the progress of academic development. Obviously, the application of opinion retrieval and opinion mining technologies can effectively promote the establishment and implementation of academic priority system.

### 3 Applying Opinion Retrieval and Opinion Mining in Scientific Research and Academic Appraisal

Currently, opinion retrieval and opinion mining are mainly targeted to non-academic opinions such as product comments and public comments. Of course, the technologies to process these opinions can be used as a basis to process academic opinions. Additionally, we should develop more efficient information retrieval and mining technologies based on the features of academic opinions. We can call them opinion retrieval and opinion mining targeting scientific research and academic appraisal. For scientific research, standard academic researches need to review literature while the main content of literature review is to summarize academic opinions related to research topics. Therefore, in response to this feature, we can collect academic opinions under different research topics into a database sorted by time sequence to provide basic references for certain research topics (Liu, 2004). On the one hand, it helps researchers to quickly know others' academic opinions and contributions. It also saves time of literature search and study. Thus, more efforts can be put on thinking and exploration. On the other hand, the novel academic opinions can be known and used by more peers as a basis in the next steps of researches in a timely manner. All these will actively promote the academic development.

For academic appraisal, related research and preparation are needed to implement opinion search and conclusion search. For example, academic databases of academic opinions with bilingual texts (Chinese and English) are needed. These databases can collect all commonly used opinions in various disciplines of academia in the world for searching and translation. Thereby, "the World's No. 1" rough translation can be avoided. Such phenomena were quite common. Some search institutions deliberately or accidentally translated the "innovation" texts to prove the novelty that "similar or the same reports to this item have not been found". It severely damaged the search authority and academic value.

To conclude, the application and development of opinion retrieval and opinion mining will bring about a series of transformation in academic research and academic appraisal and trigger scientific and talent revolution (Liu, 2008b). Implementing substantial search (opinion search / conclusion search) alone can vastly reduce waste arising from a huge amount of research expenses of repetitive researches and covert repetitive researches to achieve twice results with half effort (actually, saving more than 50% of research expenses), improve output-input ratio of scientific research expenses and optimize the resource allocation. Of course, to realize the effective application in scientific research and academic appraisal, technical experts of opinion retrieval and mining need more efforts to conduct related researches and preparation. The cooperation with technology management experts and soft science experts may be necessary and effective.

## References

1. Xu, X., Lin, S.: Opinion Retrieval Technology Research for Chinese Texts. Fuzhou University Journey (Natural Science Version) (5) (2010)

2. Yao, T., et al.: Literature Review of Text Opinion Mining. *Chinese Information Journal* (5) (2008)
3. Huang, P., et al.: A Discussion about Academic Quantification. *Dushu* (2) (2002)
4. Liu, Y.: The Intelligence Industry Revolution: Scientific and Technological Transformation, Industrial Transformation, and Social Transformation under the Coercion of Irreversible Growth of Ruin-causing Knowledge. In: Chinese with English Abstract. Contemporary China Publishing House, Beijing (2007)
5. Liu, Y.: On a complex method of evaluation which will overcome peer review. Chinese with English Abstract, *Studies in Dialectics of Nature* 20(1), 98–102 (2004) ISSN 1000-8934
6. Liu, Y.: A Galileo-Style Revolution—Talent Revolution and Science Revolution Triggered by Evaluation Method through Display of Innovative Points and Talent Recognition System through Novelty-Inquiry Analysis. *Time Education* (5) (2008)
7. Liu, Y.: E-Academy: The Next Scientific Revolution. Chinese with English Abstract, *Cross Century* (3) (2008); Address: No.55 Zhongguancun East Road, Beijing, China 100190  
Author and Affiliation: Liu Yidong, Chinese Academy of Sciences

# Author Index

- Babaie, Shahram 177  
Bannatyne, Ross 387  
Bao, Fumin 913  
Bi, Mingxue 935  
Bo, Li 547  
Bu, Xian-de 781
- Cai, Jin 329  
Chao, Yun 127  
Che, Zihui 875  
Chen, Baixiao 457  
Chen, Chang 667  
Chen, Chunxiang 845  
Chen, Gengbiao 105, 111  
Chen, Lin 71  
Chen, Ruimin 687  
Chen, Tzu-Wei 823  
Chen, Wenjuan 705  
Chen, Xiaomin 727  
Chen, Xu-ping 953  
Chen, Yading 61  
Chen, Youjun 305  
Chen, Yu 527  
Chen, Yuefen 637  
Cheng, GuangLin 759  
Cheng, Xinzhou 541  
Chi, LeJun 647  
Chu, Xuezheng 277  
Crisan, Marius 605  
Cui, Jianzhu 41  
Cui, Zheng 427
- Dai, Chien Yun 823  
Dang, Sanlei 687  
Deng, Ren 401
- Deng, Song 787  
Deng, Wenxin 89  
Ding, Shiyong 535  
Ding, Ying 1011  
Du, Changhua 993
- Fahn, Chin-Shyurng 269  
Fan, You-hong 401  
Feng, Shi-lun 371  
Feng, Yongmao 137  
Fu, Ming-gang 667  
Fu, Tingbo 255
- Gao, Haolin 713  
Gao, Xiaorong 99  
Gao, Xiuhui 597  
Gong, Jian 41  
Gong, Shangfu 33  
Gong, Weimin 773  
Gong, Xingyu 33  
Guan, Jian 541  
Guo, Jianqiang 99  
Guo, Jinghong 629  
Guo, Jun 983  
Guo, Shu-hang 899
- Han, Hong 973  
Han, Su 127  
Han, Weijuan 853  
Han, Yong 371  
Hao, Tinglei 923  
Harms, Christian 747  
He, Feng 861  
He, Guojin 845  
He, Hongying 305

- He, Junjie 637  
 He, Tangmei 355, 363  
 He, Yongqi 291  
 Hou, Yingkun 867  
 Hu, Huaping 71  
 Hu, Shaoxing 165  
 Hu, Xiangyi 379  
 Hu, Zelin 155  
 Huang, Hao 983  
 Huang, Li Ping 695
- Jia, Jing 875  
 Jia, Zhuo-sheng 19  
 Jiang, Chengzhi 807  
 Jiang, Guixiu 965  
 Jiang, Jiulei 861  
 Jiao, Peng 83  
 Jin, Yao 673  
 Ju, Yifeng 441
- Kang, Fengju 973  
 Kang, Xiaoning 815  
 Kao, Chang-Yi 269  
 Khadem, Arash 177  
 Köster, Frank 747
- Lee, Seon Ung 801  
 Li, Bicheng 713  
 Li, Binlin 347  
 Li, Changzheng 913  
 Li, Chuntian 993  
 Li, Da-peng 765  
 Li, Dayun 899  
 Li, Hao 219, 225  
 Li, Hongyu 61  
 Li, Jingang 837  
 Li, Jinlong 99  
 Li, Jun 371  
 Li, Junfei 727  
 Li, Li-xin 417  
 Li, Pu-lin 371  
 Li, Qiang 51  
 Li, Qianmu 191  
 Li, Shan 659  
 Li, Shaoqian 61  
 Li, Weiwei 787  
 Li, Xiao-ying 569  
 Li, Xiyao 493  
 Li, Xueen 853  
 Li, Ying 557
- Li, Zhaohui 277  
 Li, Zhengbin 291  
 Li, Zhenjun 83  
 Li, Zhipeng 41  
 Liang, Shaoquan 923  
 Liang, Ye 615  
 Liao, Chia-Wen 1003  
 Lin, Hong 501  
 Lin, Weimin 245, 787  
 Lin, Yong 205  
 Lin, Zezhao 923  
 Liu, Bo 71  
 Liu, Fuqiang 41  
 Liu, Gang 679  
 Liu, Hao 137  
 Liu, Hongbin 629  
 Liu, Jianbo 233  
 Liu, Jian-ping 569  
 Liu, Lan 261  
 Liu, Lei 705  
 Liu, Pengyuan 737  
 Liu, Shibin 837  
 Liu, Wei 837  
 Liu, Xiuluo 891  
 Liu, Yan 759  
 Liu, Yaxing 629  
 Liu, Yi-Bing 341  
 Liu, Yidong 1019  
 Liu, Yu 379  
 Liu, ZongRen 647  
 Liu, Zuo-Zhi 409  
 Lo, Edward W.C. 205  
 Lorenz, Tobias 747  
 Lu, Fang 89, 765  
 Lu, Jianye 427  
 Lu, Junfeng 321  
 Lu, Wen-qiang 321  
 Lu, Yu 427  
 Luan, Yuncai 867  
 Luo, Chun-yan 569  
 Luo, Yi 993  
 Lv, Changzhi 773
- Ma, Aiwen 947  
 Ma, Bin 493  
 Ma, Long 737  
 Ma, Zhanguo 379  
 Mao, Qian 329  
 Meng, Li 127  
 Mi, Shuangshan 737

- Miao, Changyun 137  
 Min, Lian-quan 51  
 Moon, Il-Young 801  
  
 Nagaoka, Hiroshi 117  
 Nesimi, Ertugrul 557  
  
 Pang, Alan 387  
 Peng, Duo 393  
 Peng, Xiao 947  
 Peng, Yongzhen 493  
 Ping, Jin Song 477  
  
 Qi, Chuanda 637  
 Qi, Lanlan 527  
 Qi, Mei 829  
 Qi, Yong 191  
 Qin, Guodong 457  
 Qin, Lu 329  
  
 Rau, Dar-Chin 823  
 Rong, Feng 137  
 Ruan, Jiangjun 433  
  
 Saravanan, R. 1  
 Shang, Kun 477  
 Shen, Zuochun 427  
 Shi, Cong-cong 245  
 Shi, Huixuan 277  
 Shi, Jian 787  
 Shu, Jie 965  
 Song, Chunlin 41  
 Song, Dianbin 347  
 Song, Qizhu 947  
 Sun, Hao 773  
 Sun, Jiaomei 255  
 Sun, Ningping 117  
 Sun, Qingjie 705  
 Sun, Yajuan 501  
  
 Tan, Guoxin 923  
 Tang, Jun-jun 297  
 Tao, Quan 347  
 Tao, Zhiyong 329  
 Thielen, Daniel 747  
 Tsai, Jia-Liang 269  
  
 Vivekananth, P. 1  
  
 Wang, Bo 983  
 Wang, Caiyi 111  
  
 Wang, Chen 787  
 Wang, Dayin 577  
 Wang, Deng-feng 765  
 Wang, Fang 19  
 Wang, Hanliang 433  
 Wang, Hong 335  
 Wang, Jianbiao 393  
 Wang, Jiangjing 861  
 Wang, Jue 417  
 Wang, Junfeng 947  
 Wang, Li 99  
 Wang, Minjie 27  
 Wang, Na 477  
 Wang, Ping 629  
 Wang, Shuli 535  
 Wang, Shuying 493  
 Wang, S.J. 931  
 Wang, Tianzuo 71  
 Wang, Tikui 99  
 Wang, Wei 597, 891  
 Wang, Xia 261  
 Wang, Xiaodan 255  
 Wang, Xin 773  
 Wang, Yan 191  
 Wang, Yuan 33  
 Wang, Yuchen 61  
 Wang, Yunsheng 983  
 Wang, Zhen 477  
 Wang, Zhong-jian 781  
 Wei, Fusheng 687  
 Wei, Jiang 485  
 Wen, Jiangtao 527  
 Wen, Qiaoyan 577  
 Wu, Canlong 185  
 Wu, Erbing 165  
 Wu, Feng 891  
 Wu, Fengjie 355, 363  
 Wu, Liming 355  
 Wu, Xian-liang 667  
 Wu, Xiaoyan 875  
 Wu, Zhifeng 965  
 Wu, Zhiyong 527  
  
 Xia, Mingchao 205  
 Xia, Yunchao 883  
 Xiao, Yong 687  
 Xiong, De-min 953  
 Xu, Huibin 993  
 Xu, Jia 935  
 Xu, Jian 867



- Xu, Jin 577  
 Xu, Shu-bin 11  
 Xu, Zhanyang 695  
 Xue, Q.Z. 931
- Yan, Guangrong 587  
 Yan, Qin 659  
 Yang, Cheng 233  
 Yang, Fan 569  
 Yang, Haiyong 335  
 Yang, Jianzheng 883  
 Yang, Jinfeng 687  
 Yang, Qing 493  
 Yang, Ruike 815  
 Yang, Yang 105  
 Yang, Yueguang 433, 441  
 Yao, Linhai 973  
 Ye, Nan 477  
 Ying, Li 547  
 You, Wen 347  
 Yu, Gengshen 795  
 Yu, Guang 255  
 Yu, Kang 127  
 Yu, Li 509, 519  
 Yu, Ligeng 853  
 Yu, Shiyun 219, 225  
 Yu, Yingda 637  
 Yu, Yong 245, 807  
 Yu, Yongtao 1011  
 Yun, Wang 623
- Zeng, Qingke 335  
 Zeng, Xinhua 155  
 Zeng, Yunbing 305  
 Zeng, Yuqin 587  
 Zhan, Shuying 185  
 Zhang, Aiwu 165
- Zhang, Bo 807  
 Zhang, Hao 629  
 Zhang, Huijuan 329  
 Zhang, Jian 155, 441  
 Zhang, Jianqi 913  
 Zhang, Jing 71  
 Zhang, Junchi 947  
 Zhang, Peng 829  
 Zhang, Rui 409  
 Zhang, Shujun 493  
 Zhang, Tao 245, 787  
 Zhang, Xianyong 965  
 Zhang, Xiaobo 815  
 Zhang, Xiaojian 807  
 Zhang, Xiaomei 845  
 Zhang, Xin-lun 781  
 Zhang, Yunning 673, 679  
 Zhang, Zhenrong 291  
 Zhang, Zhong-xiang 667  
 Zhang, Zi Hua 145  
 Zhao, Gang 587, 597  
 Zhao, Sufen 313  
 Zhao, Yongwei 713  
 Zheng, Dabin 467  
 Zheng, Shouguo 155  
 Zhou, Bing 527  
 Zhou, Jun 433, 441  
 Zhou, Sao-qi 401  
 Zhou, Xue 509, 519  
 Zhou, Zhengjuan 659
- Zhu, Lianxuan 27  
 Zhu, Xiangcai 867  
 Zhu, Xian-zhen 51  
 Zhu, Xiaohuan 795  
 Zhu, Yu-jia 781  
 Zhu, Zede 155